DOCUMENT RESUME

ED 411 322 TM 027 391

TITLE	Making Education Count: Developing and Using International Indicators.
INSTITUTION	Organisation for Economic Cooperation and Development, Paris (France). Centre for Educational Research and Innovation.
ISBN	ISBN-92-64-14050-6
PUB DATE	1994-00-00
NOTE	368p.
AVAILABLE FROM	OECD Washington Center, 2001 L Street, N.W., Suite 650, Washington, DC 20036-4922; phone 1-800-456-6323 ($58, plus $4.50 shipping and handling).
PUB TYPE	Collected Works - General (020)
EDRS PRICE	MF01/PC15 Plus Postage.
DESCRIPTORS	Academic Achievement; Context Effect; Educational History; Educational Policy; *Elementary Secondary Education; Evaluation Methods; Foreign Countries; Higher Education; *International Education; International Studies; *Outcomes of Education; *Policy Formation; Political Influences; Public Policy
IDENTIFIERS	*Educational Indicators; Organisation for Economic Cooperation Development

ABSTRACT

The chapters in this book were prepared during the second phase of a study conducted by the Organisation for Economic Cooperation and Development to develop a set of International Indicators of Education Systems (INES). They were among the many studies presented at the General Assembly of the INES Project in Lugano (Switzerland), September 16-18, 1991. Taken together, these papers present what is currently known about the organization, development, measurement, and uses of international education indicators. Attention is given to the political contexts within which education indicators are used for informing policy-makers. The 18 chapters deal mainly with conceptual and analytical issues in the organization of education indicators. They are grouped thematically into four parts. Part I, "History and Definition of Indicators," presents a framework for the other contributions. Part II, "Development of Indicators," discusses issues in the development and implementation of different types of indicators. Part III, "Indicators of Outcomes of Education," is concerned with indicators of learning, student achievement, and other educational outcomes, such as labor market destinations. The five chapters in Part IV, "Interpreting Indicators for Policy," focus on the uses and abuses of reporting and interpreting international education indicators. Each chapter contains references. (Contains 10 tables and 16 figures.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

Centre for Educati and Innovation

ED 411 322

Making Ed ati
Co

Developi nd U
Internatio Indic

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
IN OTHER THAN PAPER COPY
HAS BEEN GRANTED BY

J.R. WEST

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it
- ☐ Minor changes have been made to improve reproduction quality
- ● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

TM027391

ECD

BEST COPY AVAILABLE

Centre for Educational Research and Innovation

Making Education Count

Developing and Using International Indicators

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Pursuant to Article 1 of the Convention signed in Paris on 14th December 1960, and which came into force on 30th September 1961, the Organisation for Economic Co-operation and Development (OECD) shall promote policies designed:

- to achieve the highest sustainable economic growth and employment and a rising standard of living in Member countries, while maintaining financial stability, and thus to contribute to the development of the world economy;
- to contribute to sound economic expansion in Member as well as non-member countries in the process of economic development; and
- to contribute to the expansion of world trade on a multilateral, non-discriminatory basis in accordance with international obligations.

The original Member countries of the OECD are Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The following countries became Members subsequently through accession at the dates indicated hereafter: Japan (28th April 1964), Finland (28th January 1969), Australia (7th June 1971), New Zealand (29th May 1973) and Mexico (18th May 1994). The Commission of the European Communities takes part in the work of the OECD (Article 13 of the OECD Convention).

The Centre for Educational Research and Innovation was created in June 1968 by the Council of the Organisation for Economic Co-operation and Development and all Member countries of the OECD are participants.

The main objectives of the Centre are as follows:

- *to promote and support the development of research activities in education and undertake such research activities where appropriate;*
- *to promote and support pilot experiments with a view to introducing and testing innovations in the educational system;*
- *to promote the development of co-operation between Member countries in the field of educational research and innovation.*

The Centre functions within the Organisation for Economic Co-operation and Development in accordance with the decisions of the Council of the Organisation, under the authority of the Secretary-General. It is supervised by a Governing Board composed of one national expert in its field of competence from each of the countries participating in its programme of work.

Publié en français sous le titre :
ÉVALUER L'ENSEIGNEMENT
De l'utilité des indicateurs internationaux

Reprinted 1994

© OECD 1994
Applications for permission to reproduce or translate all or part of this
publication should be made to:
Head of Publications Service, OECD
2, rue André-Pascal, 75775 PARIS CEDEX 16, France

Foreword

The chapters in this volume were prepared by consultants and the CERI Secretariat during the second phase of the OECD study designed to develop a set of International Indicators of Education Systems (INES). They were among the many studies presented at the General Assembly of the INES Project held in Lugano, Switzerland, 16-18 September 1991 at which all the participants in the Project met to take stock of the achievements and the options for work to be undertaken during the third phase.

Taken together, these papers present what is currently known about the organisation, development, measurement and uses of international education indicators. Much attention is given to the political contexts within which education indicators are used for informing policy-makers. The eighteen chapters deal mainly with conceptual and analytical issues in the organisation of education indicators. They are grouped thematically into four parts. Part I presents a framework for the other contributions. Part II discuss issues in the development and implementation of different types of indicators. Part III is concerned with indicators of learning, student achievement and other educational outcomes, such as labour market destinations. The five chapters in Part IV focus on the uses and abuses of reporting and interpreting international education indicators.

While voicing a number of cautionary notes on the difficulties besetting indicator construction and interpretation, the distinguished authors of the various chapters are generally supportive of the OECD initiative to develop a coherent set of reliable and policy-sensitive education indicators. Their support was an essential factor in the decision to publish the thirty-six indicators that resulted from the second phase of the INES Project. That publication, *Education at a Glance: OECD Indicators*, was released on 24 September 1992, one year almost to the day after country delegates and experts attending the Lugano Assembly received the recommendation to proceed. By working with Member governments and invited experts, the OECD has demonstrated in *Education at a Glance* that it is possible to produce reliable indicators of education systems. Moreover, the interest taken by policy-makers – as well as the broad press coverage given to the publication – shows that the OECD indicators succeeded in addressing the demand for internationally comparative information on education. Whereas the scope of the first edition of the indicators report is still limited, future editions will expand the range of indicators and improve their reliability and comparability.

This volume was prepared by Albert Tuijnman and Norberto Bottani of the CERI Secretariat and is published on the responsibility of the Secretary-General of the OECD.

Table of contents

Introductory address *by* ***Thomas J. Alexander*** 13

Part One

HISTORY AND DEFINITION OF INDICATORS

Chapter 1

International education indicators: framework, development and interpretation

Norberto Bottani *and* ***Albert Tuijnman***

History and definition of indicators 21

  History of a failure 21

  Policy priorities and changing research agendas 22

  Call for a new concept of accountability 23

  Education indicators back on the agenda 25

Development of indicators 26

  Indicators in a political context 26

  Issues in the measurement of education 27

  Specificity in a comparative context 28

  The management of indicator development 30

Interpreting indicators for policy 30

  Organisation of the indicators 31

  Relationships among the indicators 32

Conclusion 33

Bibliography 34

Chapter 2

Observations on the structure, interpretation and use of education indicator systems

Anthony Bryk *and* ***Kim Hermanson***

New interest in education indicators 37

Basic premises concerning schools and schooling 38

  Schools as organisations 38

  Multiple, diverse and sometimes conflicting goals 39

  The systemic nature of schooling processes 40

Control over schooling 41

An uncertain connection: indicators and educational improvement 42

  Limitations imposed by social science 43

The limits of external control through indicator data 44
Implications for indicator systems 45
Prudent aspirations 45
An enlightenment function 45
A specific objective: enriching public discussion 46
Structuring the system: an information pyramid 47
Interpreting and reporting statistical results 47
Standards of comparison 48
Types of statistics reported 48
The indicator system in the broader social framework 49
Nurturing the technical and information user communities 49
The indicator agency and the public trust 49
Conclusion 50
Bibliography 52

Chapter 3

Knowledge, ignorance and epistemic utility: issues in the construction of indicator systems

Carla Fasano

Education indicators in the research literature 55
Education indicators as a contemporary phenomenon 56
Defining the focus on information and knowledge 56
Indicators and paradigm development in education 57
An early undifferentiated discipline area 57
From the late 1800s to the late 1940s 58
From the late 1950s to the late 1980s 59
Paradigms reflected in current indicator systems 62
Maximising knowledge 63
Issues of validity 63
Issues of coverage 64
The focus on schools 65
Managing ignorance 68
Conclusion 71
Notes 73
Bibliography 75

Chapter 4

Choosing indicators

Desmond Nuttall

What are indicators? 79
Indicators in the policy-making process 81
How are indicators chosen? 82
Policy considerations 82
The modelling approach 84
Scientific and technical issues 85
Practical considerations 88
Who chooses the indicators? 88
Criteria for choosing, developing and evaluating indicators 89
Conclusion 93
Bibliography 94

Part Two

DEVELOPMENT OF INDICATORS

Chapter 5

Education indicators: a review of the literature

Tim Wyatt

Antecedents of education indicators 99
Trends in the literature 102
The rationale for using indicators 103
What are education indicators? 104
Indicator systems and models 106
Potential uses for indicators 108
Indicators in action 110
Developing improved indicators 112
Conclusion 115
Bibliography 116

Chapter 6

Developing and implementing education indicators

Rolf K. Blank

Issues in developing education indicators 123
  Comparability and data specification 124
  Commitment, cost and follow-through 126
  Analysis and interpretation of data 127
  Data procedures and reporting standards 128
Strategies for developing a set of education indicators 128
  A consensus approach to solving problems 128
  Ensure support by demonstrating the uses of indicators 128
  Project management and country representation 129
Twelve steps in constructing indicators 130
Conclusion 133
Bibliography 135

Chapter 7

The importance of opinion in studying the education system

Robert Ballion

The actors in the system 137
Surveys of opinion 138
Opinion polls on education in France 138
Some recurring themes 140
Conclusion 142
Bibliography 143

BEST COPY AVAILABLE

Chapter 8
Process indicators for school improvement
David Hopkins

Definitions 145
    Performance and process indicators 145
    School effectiveness and improvement 146
    What is process? 150
    Summary 154
Review of process indicators 155
    Practices, policies, and organisation 155
    The culture of the school 158
    Strategies 159
    The external perspective 161
    Summary 164
Applications of process indicators 165
    International perspectives 165
    Using process indicators 166
Conclusion 166
Bibliography 168

Part Three
INDICATORS OF OUTCOMES OF EDUCATION

Chapter 9
Evaluating school performance in France
Martine Le Guen

Why student evaluation has become more important 173
The nature of school evaluation in France 175
    National assessment of student learning 176
    Participation is necessary for successful evaluation 177
    Sound methods ensure reliable findings 177
    Disaggregation 179
Three sectors of student evaluation 180
    The monitoring of student achievement 180
    Interdependence of monitoring and policy evaluation 186
    Improved understanding of the system should improve learning 188
Conclusion 191
Annex 9.1: Comparison of the international standard to the French primary/secondary . . . 194
Annex 9.2: Evaluations in French schools 195
Bibliography 196

Chapter 10
Theories of mental development and assessing educational outcomes
Ivan Ivic

Scope and purpose of the chapter 198

- Vygotsky and the assessment of educational outcomes 199
  - Allomorphic development 200
  - Mental development and school learning 201
  - Metacognitive abilities 205
- Piaget and indicators of educational outcomes 207
  - Formation of the system of intellectual operations 207
  - Cognitive structures and the content of knowledge 209
  - Achievement, understanding and metacognition 211
- Tentative conclusions 212
  - Critical analysis of the existing systems of assessment 213
  - Interpretation of assessed educational outcomes 214
  - Conceptual bases for new models of assessment 214
- Bibliography 217

Chapter 11

The implications of cognitive psychology for measuring student achievement

Herbert J. Walberg** and **Geneva D. Haertel

- The impact of cognitive psychology on the measurement of intellectual abilities: a selective history 220
  - Models of human ability 220
  - A taxonomy of human abilities 221
  - Multiple intelligences: a new perspective 221
  - A developmental perspective 222
  - Models of information processing 222
  - Implications for the assessment of school achievement 223
- The impact of cognitive psychology on the measurement of school achievement 224
  - Declarative and procedural knowledge, and personal theories 224
  - The critical role of prior knowledge 226
  - Alternative routes to achievement 229
  - New directions in research on assessing achievement 231
- Conclusion 232
- Bibliography 233

Chapter 12

Achievements of learning

Patricia Broadfoot

- Limitations of traditional assessment 238
  - Criticisms of traditional performance testing 238
  - The "backwash" effect of testing 238
  - The link between theory and practice 239
- Attempts to "map" learning outcomes 240
  - Curriculum theory-based models 240
  - Pragmatic definitions of competence 241
  - Global definitions – "personal effectiveness and expertise" 242
- New approaches to assessment 243
  - Defining a framework for recording competencies 244
  - Techniques of assessment 248
  - Implementing performance assessment 248
- International studies of student achievement 252

Assessment in the United States 253
Assessment in England and Wales 253
Assessment in Australia, France and New Zealand 254
International surveys of student achievement 255
Problems in conducting international achievement studies 255
Technical problems 255
Problems of application 257
Conclusion 258
Bibliography 260

Chapter 13

Labour market outcomes as indicators of educational performance

Russell W. Rumberger

Education and the economy 265
Rationale for focusing on economic outcomes 267
Pecuniary outcomes 267
Other economic outcomes of education 269
Difficulties in assessing the economic benefits of education 270
The role of education 271
The functioning of the labour market 272
The impact of education on productivity and earnings 273
Implications 274
A conceptual model for developing labour market indicators 275
Educational outputs 276
Labour market outcomes 278
Social and political outcomes 279
Development of economic indicators of education systems 279
Conclusion 282
Bibliography 283

Part Four

INTERPRETING INDICATORS FOR POLICY

Chapter 14

Indicators and accountability at the school and classroom level

Michael Eraut

Transactions with an accountability dimension 289
The role of indicators in such transactions 290
Structures of responsibility 291
The teacher as a reflective professional 292
The professionalisation of school management 294
The role of indicators in classroom decision-making 296
Decisions at classroom level 297
Decisions at individual level 297
Four approaches to decision-making 298
Problems in decision-making 299
The role of indicators in internal accountability 300

The role of indicators in external accountability 302
Conclusion 305
Bibliography 306

Chapter 15

Interpretation and analysis of cost-per-pupil indicators in international comparisons of education systems

Nico P. Mol** and **Frans Kaiser

Coherence among indicators 307
Heuristic approach to interpretation 308
A review of the literature 309
    Economies of scale 309
    Effort spent on education 310
    Development of an interpretation scheme 310
Interpretation of the complex variable "costs per pupil" 311
Analysis of cost-per-pupil variances 312
Examination of cost-per-pupil comparisons 318
Conclusion 320
Bibliography 321

Chapter 16

National education indicators and traditions of accountability

David K. Cohen** and **James P. Spillane

Design of indicator systems 323
The structure of school systems 326
Traditions of political responsibility 328
Professionalism and political culture 331
    Professionalism and public service 331
    Craft knowledge establishments 332
Indicators and knowledge use 334
Conclusion 334
Bibliography 336

Chapter 17

Indicators, reporting and rationality: understanding the phenomena

Alan Ruby

Information and decision-making in education 339
The demand for information 341
Some country examples 345
What happens to the information: who uses it? 348
A celebration of rationality? 350
Conclusion 352
Bibliography 353

Chapter 18

Policy uses and indicators

Linda Darling-Hammond

The relationship of education indicators to policy 357
  National goals for education 358
  Curriculum goals 359
  Policy functions of indicators 360
The potential of indicators for changing policy and practice 361
  Understanding what an indicator value means 361
  Understanding relationships among indicators 362
  Attributing causality 362
  Drawing appropriate policy inferences 363
Dangers in over-using indicator data for policy development 365
The burden that policy use places on performance measures 367
Using indicators to replace decision-making: example of misuse 368
Using indicators to inform decision-making: example of wise use 371
Conclusion 373
Bibliography 376

List of tables

4.1 Three indicators derived from the research on school effectiveness 91
6.1 Three options for collecting data for international education indicators 125
7.1 Overview of education polls conducted in France from 1973 to 1988 139
7.2 Agencies that initiated education polls in France, and their principal target groups (polls 1988-1990) 140
7.3 Public attitudes towards goals for secondary education 141
11.1 Differences in the ways expert and novice performers organise and monitor problem-solving tasks 228
12.1 An example of descriptive reporting 245
12.2 An example of assessment specifications 246
13.1 Relative wages of workers by education level in the United States, 1963-1986 268
13.2 Economic outcomes associated with education 269

List of figures

3.1 Taxonomy of ignorance 70
4.1 Linking elements of the education system 86
4.2 General framework for the definition of international indicators of education systems 87
8.1 A holistic approach to school improvement 153
12.1 Model illustrating the relationship between emerging competence and contextual understanding 243
12.2 The experiential learning cycle 251
13.1 General framework for the definition of education indicators 276
13.2 Conceptual model linking educational outputs to labour market outcomes 277
15.1 General framework for the definition of international indicators of education systems 308
15.2 Interpretation scheme for the cost-per-pupil indicator 312
15.3 Analytical framework: the co-ordinate system 313
15.4 Analytical framework: the reference value 314
15.5 Analytical framework: depicting variances in costs 314
15.6 Analytical framework: identification of counterparts 315
15.7 Analytical framework: identification of variances 316
15.8 Analytical framework: assessing explanatory weights 316

Introductory Address*

by

Thomas J. Alexander
Director for Education, Employment, Labour and Social Affairs
Organisation for Economic Co-operation and Development

It is both a pleasure and honour for me to address this General Assembly at its opening session. 1991 marks the 700th anniversary of the Swiss Confederation and we all realise that this is an important moment in the history of this country. I think it is a happy coincidence that had us schedule this meeting to take part, as it were, in the celebration. We might also take this as a reflection on the time-scale of human enterprises, although I would not suggest that we need to take such a long-term view of our common work.

I would like first of all to thank the Swiss authorities who put so much work and effort in getting us together today in such pleasant surroundings. We at the Secretariat of the OECD know from first-hand experience how difficult it is to organise big meetings and ensure that once started they run smoothly in spite of the usual last-minute problems. Our Swiss hosts have provided us with work-conducive arrangements – and this includes the special evenings they have in store for us – and it is now up to us all to ensure that these are as effective as the organising team has clearly set about making them.

I very much appreciate the effort that Switzerland and other countries put into working with us to create a lasting interaction, since long-term support and involvement are the key factors for making progress in developing educational statistics and indicators and thereby improving the comparative knowledge base of education systems.

Many of you are not newcomers to OECD's work aimed at the development of education indicators, and indeed took an active part in the various meetings and events that have taken us from a modest beginning to this more developed stage. Some of us, though, and this includes me, are not old-timers and it might be worthwhile taking a couple of minutes to review the developments that brought us here today.

The INES Project on International Indicators of Education Systems was launched at the OECD because decision-makers in key countries felt the need to get together to discuss the instruments that would help them better assess the effectiveness of their education systems and monitor their evolution. The question though, is not simply an instrumental one, and the OECD sponsored two international conferences designed to clarify the issues involved. The first, hosted by the U.S. Government in Washington, D.C. in November 1987, focused on the role of performance indicators; while the second, initiated by the French authorities in March 1988 and held in Poitiers, explored the role of indicators in evaluation.

---

* This is an abridged version of the opening address delivered by Mr. T.J. Alexander at the second General Assembly of the INES Project on International Education Indicators, held in Lugano-Cadro, Switzerland, on 26-27 September 1991.

On the basis of these conferences and the issues they raised, a research project was approved by the Governing Board of OECD's Centre for Educational Research and Innovation (CERI) in May 1988. The exploratory work undertaken to demonstrate the interest in and feasibility of education indicators was reviewed a year and a half later in Semmering at the first General Assembly of project participants hosted by the Austrian authorities. The overall positive assessment the work received led to the second phase of the INES Project, and it is the results of this phase which are before us today for discussion and appraisal.

The INES Project has now reached the stage at which it is able to report on a preliminary set of indicators and to comment on them. This is in line with the mandate the Governing Board of CERI assigned this second phase, but it nevertheless proved a challenging task. That participants were able to achieve this, albeit to a still limited extent, testifies to the co-operative spirit all brought to the work and to the sustained involvement of Member countries and their representatives at all levels. As initial reluctance was overcome and widespread acceptance gained, developments testify to the fact that a new approach now permeates the educational debate; this is too significant not to record as it is likely to herald major changes in the trends and issues that will drive thinking and reform movements over the next decade.

Indeed, the indicator set is not simply the result of Phase 2, but in a slightly longer perspective the product of almost four years of continued effort and pressing for further developments and better instruments. These four years bear witness to a continued high level of interest on the part of Member governments, and particularly those countries which sponsored the international conferences – the United States, France, Austria and now Switzerland; those which undertook responsibility for the leadership of a network – the Netherlands, the United States, Austria, France; or those which were actively involved in all or most of the working groups. The networks, the technical groups and the Secretariat acknowledged this involvement and strove to respond by focusing on the policy relevance of the indicators.

The preliminary set of indicators presented in the draft of the 1992 publication *Education at a Glance* is before us for review and comment, along with the other products of the various activities undertaken. These include a handbook prepared in 1991 by a small editorial group to bring together the descriptive templates that explain and define the indicators and to describe the organising frame that helps understand their interactions. A series of theoretical and technical papers offering specialised insights into some of the issues connected with indicators development is also presented, and each network and technical group puts forward a report illustrating the most salient points in its domain. All of these represent an impressive – and some might even think overwhelming – pile of material, but we think it important to document the work and to disseminate it, to make it available to those researchers or administrators who might want to use it.

This second General Assembly brings us together to review the work done so far, to consider its products and identify areas and issues for further work. The hasty and uninformed reader, having worked his way through the various documents and papers

I have been referring to, might jump to the conclusion that with so many caveats and qualifications raised about the use of international education indicators we might as well just forget about the whole business. I am sure that none of us here who took part in the exercise would concur – and who better than us can be aware of the many flaws, limits and frustrations of indicators?

It is because we acknowledge and accept those constraints that we are here to create collective answers to some of the questions on indicator development that are spelled out in the conceptual papers. Their tone might at times appear negative, but that is because they are exacting and demanding and penetrating. They reflect a great lucidity and a sound absence of illusion that I trust mark a coming of age in our thinking on indicators. I would say that these essays are a deeply original contribution because they force us to reason and to monitor developments carefully instead of letting our enthusiasm get the better of us. They are also a major contribution to the rest of the OECD, as in no other sector was this conceptual dimension integrated *ab initio* in the development of indicators. With the many questions and issues that we raise relevant to these sectors, this is a significant domain of expertise we have carved out for ourselves and we can be proud of that.

In many ways, *Education at a Glance* is a concrete expression of the current state of knowledge implied by the conceptual papers. This exploratory collection of figures, graphs and comments is the best that can be put together today, and it deliberately uses those papers and makes the most of their insights and observations. It is also an invitation to go ahead and progress by following the direction they point to. Together with *Education at a Glance* and the handbook, those papers shape a strategy – to identify critical issues, to have an open discussion about them, and to progress by trying out possible answers. And this is why and how present apparent weaknesses can turn into long-term strengths.

But I feel that we need to take a few minutes to consider the broad political context if we are to think about the utility of further developing education indicators in a more extensive way than by simply focusing on what is brought to this meeting. Let me, on this aspect, share my thoughts with you.

First of all, it is obvious that the political environment of OECD Member countries has changed greatly since the beginning of this work, and continues to evolve. It would have been difficult in 1987, as some of you were discussing the preliminary steps of this work, to foresee the deep-rooted changes that are taking place in the former Eastern bloc countries. These changes do not occur in a vacuum, and the OECD has recognised the need to accommodate these formidable shifts in the structure of Europe.

The Asian and Pacific region is also one in which changes have been under way, although for a longer period and with a less dramatic pace; and even though Europe has been at the forefront of political thinking, we should not forget about what the OECD calls the Dynamic Asian Economies. They bring us the challenge of different cultural and social values that might prove difficult to weave into our thinking on social and economic issues unless, again, we take this factor into account now.

Latin America is also waiting, as it were, on our doorstep with issues that are primarily those of societies facing huge economic and social imbalances and striving for democracy. We need, I believe, to take these developments into account as we think about the future shape of this work and where we want to take it, since the next few years are likely to witness further changes that we will probably want to accommodate.

Changes of a different order are also manifest – for example, the fact that widespread migration has become a long-lasting trend, one that is as germane to the world currently shaping itself as the cold war and detente were to the one many of us grew up with. It is probably not necessary to spell out how important this will be for the societies of the 21st century, but I would suggest that it is a factor for social change with a potential and a magnitude equal to that of the industrial revolution in the 19th century, or the more recent impact of new technologies on work conditions and organisation.

Equally striking is the shift in international competition which appears to be leaving the strictly economic realm to take on a broader dimension and affect society as a whole. This creates special constraints for social policies in general and education in particular. Education and training appear more and more as one of the key factors likely to influence productivity and competitiveness in the long term. While this is in line with some of the classical thinking on the economic returns of educational investment and the theory of human capital, it is a theme now imbued with a social and global urgency it did not possess before. This brings with it claims for better accountability and more effective school control which it is important to recognise, understand and monitor.

Probably the major lesson which the changes we witnessed over the past four years should bring home to us is that no nation or group of nation lives in isolation. We have lived for decades in a world organised by the international order inherited from Yalta, a clear-cut division of influence zones; and it becomes more and more apparent that – in politics, in economics, as in social or environment issues – globalisation is now the key word. This should force us to think more carefully about the impact of our actions, since we will find that more and more they affect not only ourselves or our immediate partners but also others. OECD Member countries will have to think actively about their co-operation policy and the role they wish to play. In the field of education in particular, it will be important to think through the issue of co-operation with other international organisations such as UNESCO and the EC – but I will get back to that in a minute.

At a very different level, changes are also displayed in our thinking on education and educational performance and evaluation, and I would distinguish two major trends here. Schools, pedagogy and academic excellence dominated the agenda of educational reform for many years, but it seems that the trend is now to focus on the social function and the social factors of education, and to consider schooling – the process – rather than schools in isolation. This suggests a shift towards a broader encompassing perspective on education, one that would take into account education and training, and their overall effect on skills and labour market participation, and not simply the testing results. Supporting this is the fact that the work undertaken in this specific domain in the INES project has received widespread support and produced worthwhile results in spite of the fact that this was a new area in which little work had been done during the exploratory phase.

The second major trend I would see is the recognition that the assessment of educational outputs and performance needs to be undertaken more objectively. It is understandable that a sector that deals with the transmission of culture and values be more emotionally charged than others, and societies are understandably more demanding in this case. Yet we have for years created, nurtured and conveyed a large gap between a very high level of social expectation *vis-à-vis* schools and education, and a glaringly mediocre opinion of the quality of educational services actually delivered by the system. The stakes now run high, with the risk of a destructive spiral of pessimism which might lead us into believing that no reform can put this to right – clearly a zero-sum game. And yet I feel there are signals suggesting that education authorities in OECD Member countries are aware of this dilemma. Many governments recognise the need for instruments and approaches that will enable a dispassionate analysis of the condition of education systems and which they can share with the public at large. Concepts such as accountability, evaluation and indicators explicitly and actively applied to education rank high among the signals of changing attitudes.

A third and final trend that I would like to point to is the end of educational isolationism. The general wave of educational reforms that has been apparent in most of the OECD Member countries since the 1980s and which is characterised by an overriding concern with the effectiveness of schools seems to have brought with it a new interest in comparability issues. International comparisons of educational conditions and performance are now perceived as a means of adding depth and perspective to the analysis of national situations. References to other nations' policies and results are beginning to be routinely used in discussions of education, and comparability now belongs with accountability to that changing set of driving words which shape the current management paradigm of education. Since this forces us into carefully reviewing national concerns in the light of other contexts and other solutions, it is likely to broaden the scope of actions and responses we can offer, and educational policies should be the richer for it.

What does this suggest that is relevant to our work? As I see it, we should address the need for both larger and at the same time more clearly defined subsets of collaboration, as no longer do clear-cut frontiers mark the limits of co-operation. We appear to be moving rapidly towards a world of ever-expanding and more demanding synergies, and we need to think through collectively how we want to discharge what others expect of us and to set priorities in the programmes we are asked to organise. We cannot pay lip-service to a vague ideal of co-operation, but need to put forward concrete ways of setting about implementing it.

In education and training, in many ways we have implicitly started acting on the basis of some of the trends I was just discussing. The work carried out by INES participants is an immediate outcome of the pressure for accountability, for comparability, for a broader perspective on education. We should take this further. The educational sphere is one that lends itself to a specific agenda for stronger co-operation that can accommodate the many changes under way. In particular, it is clear that the set of indicators we are reporting on reflects far more closely the political concerns of the 1980s, as expressed during the 1984 OECD Ministerial Council, than it serves those of the new decade. The Ministerial communiqué of November 1990 pointed to specific domains of current interest to Member governments, and this should guide our thinking, as can the medium-term programme of work defined for the Education Division and CERI. This will be a way of adapting and updating *Education at a Glance*, of making sure it remains attuned to present-day concerns and issues.

In this, other international organisations have a clear role, given the commonality of interest and concern often expressed; we at the OECD cannot and indeed should not discuss educational data collection or definition and comparability issues without including in our debates UNESCO and the EC as co-users and co-developers of the international questionnaire and standard classification currently used. Co-operation with UNESCO, the EC and the Council of Europe needs to take different forms appropriate to the specific concerns and mandate of each organisation, but the bottom line is the same: the drive is for enhanced educational co-operation, particularly within the broader European stage.

These reflections have taken me somewhat further than I intended from the main topic of this presentation. At the same time, I think they are relevant to our tasks of the next three days, since they address issues we should have in mind as we think about the future of this undertaking. This General Assembly is an important moment in the development of the work, and we should make sure that we are clear as to what we need to accomplish during the short period we have together. I see three goals before us, each corresponding to a different time dimension.

The first goal is past-oriented and introspective. It requires reviewing and critically appraising the work of Phase 2, commenting in particular on the material that has been disseminated for this meeting. The underlying question is: does this adequately reflect the expectations and demands that were set at the beginning of this phase? This is not an easy question, but it is one that requires an honest answer if we are to set realistic objectives for further work.

The second goal is embedded in the present and can be seen as laying the basis of an in-depth discussion on some of the conceptual and practical issues that arise from using and interpreting indicators of education. This is not an item we will be able to deal with in any definite manner during today's meeting, but rather a theme that we should come back to during the months and years to come.

The third goal looks to the future: for this endeavour to continue meaningfully, we need to identify commonalities of perspective and priorities across Member countries, and to define key lines for further work. This implies specifically discussing issues of data improvement, which include practicalities of the data-gathering process and problems of definitions and coverage, notably lack of data on flows, achievements and school processes. Another item for discussion is how the organising frame which is proposed can be refined and used. A final point I would like to mention is that of the policy relevance of the work and how we can interpret, use and disseminate the indicators to address this concern.

Last but not least, a major goal is to have you all express what you are bringing to this assembly – your expectations, your comments, your wishes. They are important because we need to reach a realistic view of what we can achieve over the next five years, given the current organisation and resource levels. This view should be guided by a clear understanding of what governments expect, what they are willing to do, and what level of effort they are able to sustain. Only thus will we be in a position to design the most efficient and appropriate mechanisms to strengthen international co-operation in a highly sensitive and complex domain.

I shall stop here, as it is high time that we moved on with our agenda and started thinking through the issues placed before us. Let me conclude by saying that although I have until now followed with great interest the progress of the various networks and technical groups, I had not had the opportunity of meeting most of you. I am pleased to be here with you today and look forward to a successful, productive meeting.

Part One

HISTORY AND DEFINITION OF INDICATORS

Chapter 1

International Education Indicators: Framework, Development and Interpretation

by

Norberto Bottani** and **Albert Tuijnman
OECD/CERI Secretariat

History and definition of indicators

History of a failure

In April 1973, the OECD issued a short document entitled: "A Framework for Educational Indicators to Guide Government Decisions". The 46 indicators described in the study were intended as measures of the effects of education on the individual and society. The indicator set, which had been prepared under the auspices of the Working Group on Educational Statistics and Indicators, was endorsed by the Education Committee of the OECD at its meeting on February 13, 1973. The organising frame for the indicators comprised six policy sectors: *i)* the contribution of education to the transmission of knowledge; *ii)* the contribution of education to achieving equality of opportunity and social mobility; *iii)* the contribution of education to meeting the needs of the economy; *iv)* the contribution of education to individual development; *v)* the contribution of education to the transmission and evolution of values; *vi)* the effective use of resources in pursuit of the above policy objectives.

It is interesting to note, about two decades *post factum*, the high expectations that surrounded the first OECD study on international education indicators, as well as the enthusiasm of the people who supported it. These high hopes reflected a position, widespread during the 1950s and 1960s, about what the social and behavioural sciences could do in improving education and, more generally, in providing a scientific and rational basis for the planning of a modern, industrial society (Bauer, 1966). However, the era of an almost naive belief in the applicability of the explanatory approach in the social sciences, which presupposes the use of the quantitative or positivist research paradigm, was rapidly coming to an end. Critical questions were increasingly asked not only about the feasibility of rational planning but also about the desirability of continuing a policy of expanding the intake capacity of education systems.

The evident gap between high expectations and promises, on the one hand, and actual performance on the other led to frustration, not least among the practitioners, and

21

sparked a heated discussion on the proper role of social and educational research. This debate also inspired a controversy over the adequacy and validity of social and education indicators, and ended the belief in a functionalist, fact-finding interpretation of the role of the social sciences in the production and application of knowledge to the solution of social and educational problems. These developments naturally had strong repercussions on the OECD study on education indicators. After an initial retreat from high ambitions, the idea to develop a set of education indicators was eventually given up entirely.

Several factors explain the quick demise of OECD's indicators project in the 1970s. Perhaps the central reason was that the study did not seek to establish a direct relationship between the indicator data and the main policy interests of that period. The indicators movement consequently failed to enlist and sustain the support of policy-makers, because it was unable to convince them that the indicators would offer unambiguous, timely and above all policy-sensitive information on the functioning of education systems. Without the interest of policy-makers and the support of national administrations and their statistical offices, the research community lacked the legitimacy it needed to continue the work, let alone embark on a major development initiative.

Policy priorities and changing research agendas

The idea of developing a set of international education indicators did not resurface until 1987. Several questions seem pertinent: Why did it take almost 15 years to overcome the doubts, and why did the OECD countries, once the interest was mounting again, not seek to revive, modify and enrich the agenda for data collection that was originally defined in the 1960s? This agenda had, after all, only been partly implemented, because the financial resources needed for realising an effective system for statistical information on education at the international level had been forthcoming only sparingly, despite some major methodological achievements in the development of comparative education. At the international level, only one apparent success was recorded during the 1970s: the international standard classification of education (ISCED) had at last reached an acceptable stage in its prolonged development and was implemented by some countries and international agencies.

The answers to the questions raised above are not self-evident. The deep feeling of disenchantment that followed the previous euphoria about the role of formal education in the building of a just, affluent and democratic society, which had characterised most of the 1960s and early 1970s, provides part of the explanation. As experimental reform programmes in education were becoming discredited and increasingly abandoned, a vast but uncompleted research agenda could no longer be sustained. Disappointment and scepticism towards educational research became a hallmark of the time, replacing the previous enthusiasm. The ensuing "paradigm controversy", moreover, had bred a sense of resignation, or even of fatalism, concerning the possibility of achieving major improvements in education through professionalism, knowledge creation and empirical, data-based research. The capacity of inert education systems to resist structural and procedural reform was emphasized, and macro-level educational planning suddenly became suspect and unfashionable among educators and social scientists, who instead turned their energies to projects designed to bring about change and innovation at the school and classroom level.

Specificity became a new catchword in educational research, and a host of innovative, qualitative studies yielded findings that were interpreted as lending support to the argument that, since each local school situation should be considered unique, aggregate comparisons are meaningless and even harmful. While contributing much useful knowledge about the factors that facilitate or inhibit school-level innovation, these micro-based studies created a tendency to overestimate the degree of dissimilarity among schools. The argument, which is frequently made, is that "schools make the difference, not the education system as such; teachers and parents account for student achievement, not the policy of decision-makers". A counter-argument is that this view risks negating the holistic perspective, since educational micro-systems, such as schools, are embedded in meso- and macro-systems.

The development of methods and procedures for the analysis of classroom interaction and school performance, which has resulted from the emphasis on micro-based studies, has greatly increased the amount of information available at the local level, thus enriching the knowledge base of teachers, principals, local education authorities and the community. The concern with local variation as a means of encouraging experimentation and solving innovation problems, combined with scepticism about the capacity of hierarchical managerial models to steer and implement reform, has provided legitimacy for the devolution of responsibility from national education authorities to municipal authorities, school principals and teachers. This devolution of decision-making power in the system is a two-edged sword, however, because, as a result, accountability also increasingly rests with the municipal authorities and the local school. The new paradigm has increased our understanding of local factors and the conditions of schooling, but it has not shed light on the functioning of the education system in its entirety.

Call for a new concept of accountability

The debate and the reform movement following the publication, in 1983, of the much publicised report, *A Nation at Risk* (United States National Commission on Excellence in Education, 1983), have had a major impact on the perception of the usefulness of aggregate data on aspects of education finance, organisation, enrolments and outcomes. The re-emergence of interest in the development of education statistics and indicators during the 1980s was a direct result of this debate. In certain countries, an increase in the level of public funding of international comparative studies in education was another consequence.

In the late 1980s and early 1990s, the publication of the results obtained in large-scale international surveys conducted by the International Association for the Evaluation of Educational Achievement (IEA) and the International Assessment of Educational Progress (IAEP) on student achievement in subjects such as mathematics and the sciences focused the attention of both the public and decision-makers on the outcomes of education. The evaluation of student and school performance, the monitoring of the functioning of education systems, the guidance of educational policies and the improvement of resources management became central issues in the political forum. Parallel to the increase in attention paid to the outcomes of education, the emphasis of politicians shifted from issues in managing the quantitative growth of the system to school improvement. Accordingly, student achievement became a key criterion for judging the quality and effectiveness of the education system.

In education a distinction is sometimes made between two models relating input and process variables to indicators of educational outcomes: models of student learning and models of educational attainment. There are two main differences between these models. The first concerns the emphasis they give to the curriculum and aspects of the classroom environment in shaping outcomes. The second has to do with the nature of the construct typically employed as the outcome variable. For example, models of educational attainment tend to use a construct associated with an educational level, assessed in terms of either years of formal education completed or an educational standard or qualification reached. By comparison, models of student learning tend to focus more on the outcomes of educational experiences in the classroom. The distinction between these two models is necessary because the processes influencing educational attainment are generally different from, but are inevitably related to, those which influence achievement measured on the basis of scores on a standardized performance test.

In the 1980s several meta-analytical reviews were compiled, mostly of research studies conducted in the United States, concerned with the factors that influence student learning. It appeared that a limited number of variables were common and essential to most studies, and that these could be traced to a few representative models, such as those developed by Carroll (1963) and Bloom (1976). Carroll's model of student learning emphasizes the time spent on the learning task (Carroll, 1989). The model features five components: *i)* student aptitude; *ii)* student ability; *iii)* perseverance; *iv)* opportunity to learn; and *v)* the quality of instruction. Modifications and extensions of this model can be found in, for example, Harnishfeger and Wiley (1976) and Hanushek (1979). In summarising the findings of several meta-analyses, Haertel *et al.* (1983) and Fraser *et al.* (1987) identified nine "generalisable" factors requiring optimalisation to increase student learning. These factors were grouped into three clusters: the environment, student aptitude, and instruction.

Research studies aiming to explain educational attainment emphasize a number of factors in addition to the micro-based variables commonly specified in the models of student learning mentioned above. Among the determinants of student attainment, relatively greater significance is attributed to the variables operating at the meso- and macro-levels of the system, such as the allocation of financial and human resources, school differentiation practices, parent-teacher interaction, partnerships with business and the community, type of school, school leadership characteristics, counselling and guidance practices, monitoring and evaluation. The hierarchical organisation of education systems, for example in terms of the distribution of decision-making capacity among the different levels, or the accountability attributed to the actors at the different levels, is considered an important variable in this perspective, because it focuses attention on structural education indicators in addition to the variables measuring aspects of the learning process. Accordingly, the linkages among the different parts and elements of the education system, and thus the consistency of the system in its entirety, are identified as central factors in explaining the level of educational attainment of a population.

In a paper written for the United States Department of Education, Kirst (1989) explored the relationship between education indicators development and accountability. He identified six accountability strategies:

i) accountability as performance testing and reporting;
ii) accountability through monitoring and enforcing compliance with standards or regulations;
iii) accountability through incentive systems;

iv) accountability through reliance on the market;
v) accountability through changing the locus of authority or control of schools;
vi) accountability through changing professional roles.

The choice of one approach over another is essentially political. Beginning in the 1980s in the United States, new strategies for improving and restructuring education have been devised in most OECD countries, and the resultant changes have promoted different models of accountability. The first four strategies have become increasingly important, but little change has occurred with respect to the locus of authority and teacher professionalism.

The important conclusion is that the responsibility for adequate performance cannot rest exclusively with the individual teacher or a single school, but has to be shared among all actors involved in the organisation and operation of the education system at all its levels. It follows that the analysis of the structural characteristics of the system – and thus the collection and reporting of aggregate data on national and international trends in education – has become an essential ingredient in an encompassing evaluation strategy. This is reinforced by a factor external to education: the development of supply-side economic policies aimed at limiting public expenditure while increasing efficiency in the public sector.

Education indicators back on the agenda

In the mid-1980s, a new scenario and new priorities emerged in the educational policies of the OECD countries. The goal became one of expanding the enrolment capacity of the system at the upper secondary and tertiary level while simultaneously improving the quality of education. Because this objective had to be achieved without adding new financial resources, the efficiency and cost-effectiveness of education also had to be improved. It was realised that these goals set high demands on policy analysis and, accordingly, that the comparative knowledge base in education was in need of improvement. It follows that, in this scenario, decision-makers attached great importance to the development of a coherent system for the monitoring and evaluation of educational progress. As the Ministers of Education of the OECD countries concluded when they met in Paris in November 1990, "information and data are preconditions for sound decision-making, prerequisites of accountability and informed policy debate" (OECD, 1992*b*). The general lack of internationally comparable information on education in the OECD countries was noted as a major drawback. This realisation led to a request addressed to the international agencies, notably the OECD, to improve the visibility, accuracy and timeliness of system-level, aggregate data on education.

In November 1987, the OECD and the United States Department of Education convened a meeting in Washington, D.C. for discussing new, feasible approaches to developing comparable statistics. The aim of the conference was to agree on a small set of indicators that could be jointly pursued by the participating countries. This implied that all OECD countries would need to reach an accord on which indicators were needed. In addition, some countries might have to consider new data collection activities.

The participants were enthusiastic about the possibility of jointly developing international education indicators but, as might have been expected at a first meeting, were unable to arrive at a concrete workplan. Nevertheless, a consensus emerged as to the strategy and the methods of work to be pursued. This would involve the voluntary

BEST COPY AVAILABLE

participation of countries in networks set up around clusters of indicators corresponding to common policy priorities. The networks and their corollary – an experts group of national correspondents from the Member countries – at a later stage proved essential for the successful implementation of a project on the development of a set of international education indicators.

Development of indicators

The development of a set of international education indicators is not merely a technical exercise planned and controlled by statisticians, but first and foremost it is a political one. An indicator is not simply a numerical expression or a composite statistic. It is intended to tell something about the performance or behaviour of an education system, and can be used to inform the stakeholders – decision-makers, teachers, students, parents and the general public. Most importantly, indicators also provide a basis for creating new visions and expectations.

Many ideological and methodological factors obstruct the development of new approaches for defining and measuring education indicators. The field of educational research may well be among the most conservative, in a methodological sense, of all the social and behavioural sciences. The lack of integration among competing theoretical and methodological paradigms in the field has created, and maintains, conservatism in the production of statistical information. Perhaps the explanation lies in the insecure nature of education as a interdisciplinary field of study.

Indicators in a political context

Indicator design involves a delicate interplay of both technical and political factors. It would be a mistake to consider that technical concerns, such as the level of data aggregation, the specification of data elements in a calculation formula, the design of a data-collection strategy, or the choice of test items, do not carry political implications. The transition from statistics to indicators is therefore a delicate passage, as the history of the economic and social indicator movements of the 1960s shows. Burstein *et al.* (1992) note that a major reason for the rapid demise of social reporting was that policy concerns were subsumed by research concerns:

> "The momentum sustaining [the legitimacy of social reporting] was short-lived. With the change of administration in 1969, the federal government began to back away from its commitment to social reporting... This retreat and the diminished financial support for social indicators has been attributed to disillusionment resulting from naive expectations and unfulfilled promises and to an overemphasis on the concerns of social science research rather than the needs of the policy community." (p. 411)

"The social indicators movement did not develop a rich collection of measures of direct policy interest or public appeal", writes Rockwell (1989) in a paper prepared for the U.S. Education Indicators Panel. The implication is that the continuation of the work on education indicators critically depends on the continued interest and involvement of the policy-makers who supported the revival of the indicators agenda in the late 1980s. Implicit in this conclusion is a peculiar view of education and of the role of the public authorities in education. First, the importance of accepting the complementarity of public and private sectors in education is underscored. Secondly, there is the view that policy-makers in a democratic society are ultimately accountable to public opinion and the users of the education system. Thirdly, the responsibility for the creation of a monitoring system for assessing educational progress and informing the policy debate rests with the State. This, again, implies that public authorities are responsible for the organisation and funding of the collection of data needed to sustain a monitoring system that offers clear and solid information about the quality of education. Thus, education indicators should be attractive not only for the statistician or the academic but for the authorities in charge of educational improvement as well as for the public at large. The conclusion is that the process of designing and implementing a set of education indicators cannot be reduced to a merely intellectual exercise, however interesting this could be for the people taking part.

The temptation is for the work to be focused on the aim of improving the scientific basis of the indicators. If this aim were to become paramount, then the publication of a set of education indicators might well be postponed indefinitely, because an indicator is, arguably, never perfect. This is not to say that one should not seek to support research on indicators. There are several needs in this respect: clarifying the conceptual framework, improving the understanding of relationships among the indicators in a set, facilitating data management, improving the adequacy of measurement tools, and developing a basis for measuring those aspects of educational processes and their outcomes for which acceptable measures have not yet become available.

Issues in the measurement of education

Several obstacles to further work on the indicators are mentioned above. Prominent among these is the lack of adequate tools for measuring the results or outputs of the education system and, even more difficult, for assessing the long-term consequences of education for the individual, the family, the local community, the workplace, and the national and global economy. The differences among the OECD countries, which are reflected in the structure and orientation of their national education systems, poses enormous challenges concerning the comparability of education data at the international level.

Traditions in educational measurement have produced a rather limited number of context variables, a large number of input measures, and few data on educational processes, student achievement and educational outcomes. The third group of variables has almost exclusively been studied by researchers in fields such as comparative education and school effectiveness. National authorities have – with only few exceptions – not included these variables in their regular data collection and reporting system. Comparable information on student achievement is still scarce, although a strategy and methodology for collecting international data on student achievement was pioneered in the 1960s, mainly by people linked with the International Association for the Evaluation of Educational Achievement (Husén, 1967). Public authorities – with the notable exception of the United States, where data from initiatives such as the Scholastic Aptitude Test and the National Assessment of Educational Progress (NAEP) have been around for decades (Haertel, 1988) – have only recently widened the scope of their data collection by including large-scale surveys of student achievement in a range of subjects.

Collectively, these initiatives have created several large databases, an impressive array of statistical methods and techniques, many useful insights into the "productive" factors in student achievement and school improvement, and much practical knowledge about the organisation and management of large international studies in education. Despite the extent of progress achieved so far, however, the data-gathering strategies adopted by the national authorities have not, in general, incorporated new possibilities for the regular collection of information on the results and outcomes of schooling. On the contrary. In most respects, the national agendas for the collection of educational data have not changed substantially over the last 30 years. This is a weakness that, by necessity, also besets any effort to establish a routine for the collection of student achievement data at an international level. Without reliable data on the outcomes of the education system, an indicator set is not only incomplete but – worse – has limited applicability. The purpose of education indicators is to inform the policy debate, assist the decision-making process and inspire policy action. Obviously, this aim cannot be achieved if only data on resource inputs and student flows are available. Outcome variables must be included in an indicator set because they offer a way of assessing the extent to which different education systems achieve their goals. Such measures of the outcomes of schooling should not be restricted to student achievement data, because all countries pursue goals in education that go beyond those concerning learning in cognitive domains.

Educational processes occur in a variety of social, cultural and economic contexts. The frame factors setting the parameters within which an education system operates differ to some extent across countries. This naturally has consequences for the meso- and micro-structures that influence the internal operations of schools and, hence, shape the teaching-learning processes in the classrooms. Differences in the contexts of education have resulted in different goals for education. Accordingly, the emphases attached to the various domains of the school curriculum differ as well. This apparent heterogeneity of contexts, goals and content has prevented the adoption of a common denominator in a definition of student achievement. The result has been widespread disagreement about what constitutes good performance and which aspects of achievement one ought to be measuring. While there is a deep consensus that an assessment of educational achievement cannot and should not be reduced to an assessment of student performance in some key subjects, it has proven to be most difficult to reach agreement on precisely which aspects one ought to be measuring in order to ensure that the multifaceted nature of education is acknowledged in all its complexity. This is not only a consequence of the real difficulties in defining and measuring such aspects, but also of an insufficiently developed theory of education. Since the output measures needed to create a meaningful set of education indicators are generally not available, educational productivity cannot be examined at the national and international levels. Without a sustained research effort, which presupposes considerable political courage and support, meaningful outcome data cannot be produced. If this were the case, then the goal of developing a coherent set of international education indicators would be unattainable.

Specificity in a comparative context

Comparability is another key issue. The term refers not only to the technical requirement of ensuring that the data are properly standardized. The difficulty of obtaining data from national authorities and of reporting them in a common format that makes them comparable is enormous but not impossible. Again, the two main obstacles are insufficient theory and inadequate knowledge about the comparative approach in education: comparing means grouping data according to some classification. Thus, knowledge of the distinctive characteristics of the systems and processes under study is at the heart of the comparative method. Moreover, classification is among the most basic of cognitive operations in systems analysis. It draws on a method that is applied for studying inanimate objects, the biological world as well as human societies. In conclusion, the comparative approach is a way to describe reality or to represent perceptions and observations. This statement, of course, begs the question of the value-added of comparative studies. What is the intrinsic cognitive value of the comparative method in education and, consequently, what are the implications for the instruments to be employed so that reliable and valid comparative knowledge is produced? The body of knowledge and insights involving comparative methodology is highly fragmented and spread across many disciplines. It needs to be organised and synthesised, so that the paradigm controversy in comparative education can be resolved.

Basically, two paradigmatic approaches are in opposition, although they could well coexist: a knowledge of similarities and one of differences. This calls attention to the parameters that make an education system "visible": the development of education indicators draws certain parameters into the limelight; others may be neglected. Accordingly, the significance attributed to given factors in education can be changed, and this in turn leads to new questions about the nature of differences and similarities among education systems, schools, classrooms and even students. This perspective was expressed by T. Neville Postlethwaite who, at the OECD Conference on Education Indicators held in Washington, D.C. in 1987, argued that the main benefits to be derived from indicator comparisons do not concern new insights into similarity but rather the identification of differences, which could then instigate further explanatory research. The nature of the observation, the methodology, the type of data to be collected and the formalism of the analytical procedures to be employed differ according to whether the comparison aims, first, to illustrate the specificity of each situation or to legitimate the peculiarities of local, regional or national solutions or, secondly, to identify the commonality of educational processes, the homogeneity of goals and objectives, and the similarities of the issues at stake.

The legitimacy of the comparative approach in education is not in doubt, but the application of various research methods, which derive from different disciplinary perspectives and conflicting views on the epistemology of science, has long been the subject of some controversy in comparative education. It can be argued that the development of international education indicators is not a neutral exercise in this respect, since it implies measurement and a high level of data aggregation. Another argument brought to bear against the initiative is that the development of international education indicators would necessarily imply a "trade-off" between cross-national "comparability" and the "fidelity" of the statistical portrait of a given country. The latter position is highly unproductive, if not untenable, because it calls into doubt the validity and applicability of the – well-established – comparative approach in the natural and social sciences.

It follows that a central aim of the OECD study on education indicators should be to improve comparability. This is a difficult task for many reasons. Not the least among these is the problem of contrasting comparisons at an international level with descriptions expressed in each country's nationally defined categories. Economists accepted this challenge almost 40 years ago, and enormous strides forward have been made since.

Before economists could embark on this endeavour, they had to come to terms with a fundamental question: What is the nature of the knowledge to be produced, descriptive or heuristic? In some disciplines this debate has been settled, but not in education. This is another mark of the limitations not only of educational theory but, more fundamentally, of the weak disciplinary and scientific status of education as a field of study. Given this weakness, it is worth remembering that the work on international indicators is not fundamental, but applied and contextualised in a political arena. As Burstein *et al.* (1992) put it:

> "Indicator development is usually driven and sustained by political interests, such as expectations that indicators can result in accurate information about the condition of education and that this information can be an integral part of the process of educational improvement. Moreover, indicators and indicator systems themselves are political entities. Their construction reflects particular assumptions about the nature and purposes of education, and they often embody beliefs about what directions reform should take. The indicators that are selected will push the education system toward the assumptions and beliefs they embody – that is, what is measured is likely to become what matters." (p. 410)

The management of indicator development

The OECD study on education indicators has developed new procedures and standards for the management of data collection and dissemination. A full account of these procedures, which have been implemented with the aim of ensuring the production of high-quality information, cannot be given within the limited scope of this chapter.

The premise behind the methodology is that the flow of high-quality data from the countries to the international organisations is not guaranteed. One cannot expect to obtain adequate data that meet comparability and other quality criteria without managing the data collection, transmission and reporting processes. Therefore, organisational resources must be devoted to all phases of a data collection activity, and active co-ordination above the national level is needed in order to allocate resources effectively and in areas where they are most needed. Thus, in order to meet information needs while safeguarding quality, managerial processes that facilitate the sound operation of an international data collection and reporting system must be put into place.

Indicators have to satisfy the information needs of legislators, policy-makers, practitioners and the public. Advisory groups on policy and methodology have been instituted during all phases of the project, partly to assist the OECD in identifying the audiences and their specific information needs. Other essential elements in the management strategy are a National Co-ordinators Group, a Technical Working Party, and several networks in which the OECD countries voluntarily combine efforts to develop and implement the steps needed for indicator construction, measurement and dissemination.

Interpreting indicators for policy

Systemic sets of indicators are in use for programme analysis and decision-making in policy sectors such as economics, public health and consumer affairs. By contrast, in the area of education, until now there has been little sophistication in using statistics as indicators. Indicators are a potentially powerful tool for defining and interpreting relationships among the different aspects of education systems. However, they need to be organised into a framework that draws attention to their interrelationships.

Organisation of the indicators

Many epistemological and practical issues arise in the construction of such a framework, and further work will be required before they are resolved. At this stage, any conceptual framework will therefore be provisional. Recognising this, in the OECD project on education indicators a pluralistic approach is taken, using both conceptual and pragmatic orientations and incorporating policy concerns. Thus, some of the indicators derive from logical relations among different parts of the education system and are empirical in nature, whereas others derive from practical concerns and are policy-sensitive in their orientation. The advantage of this combined approach is that it strikes a balance between stability and flexibility in a set of indicators that will evolve with time (van Herpen, 1992).

The OECD framework consists of three clusters of education indicators, offering information on: *i)* the demographic, economic and social contexts of education systems; *ii)* features of education systems; and *iii)* the outcomes of education. Within each cluster several indicators have been proposed and methodologies developed for measuring them. However, no indicators have yet been calculated for certain important elements of the framework, such as expectations and attitudes of the consumers of educational services, education staff characteristics, the attained curriculum, and the time the students spend learning in the classroom.

The precise nature of the linkages among the indicators is often unclear. Considered in this context, the framework so far developed serves primarily a pragmatic purpose. It provides, in simplified but systematic form, comparative information on what are widely agreed to be significant features of education systems and their contexts, and on a selection of pervasive issues which arise through their functioning and development (OECD, 1992*a*). Because many topics, however important, seem to evade straightforward quantification – examples are the curriculum, teaching methods, social equity, and non-curriculum-bound knowledge and skills – the framework is necessarily incomplete. It will be superseded in the future as understanding of education and society evolves and improvements in data sources take place.

To be useful, indicators have to satisfy certain conceptual and methodological criteria. Comparability has already been mentioned. Other important criteria are accuracy, validity and interpretability. The availability of accurate data is a self-evident precondition for indicator construction. But the production of high-quality data at the national level does not necessarily imply that sound data can be supplied at the international level. Consequently, the relationship between national data producers, data suppliers and users at the international level is an important determinant of the accuracy of the information offered by indicators. Validity, which refers to whether an indicator actually describes the phenomenon it is believed to be associated with, is not less important but certainly more difficult to establish. Interpretability is the ultimate criterion. It refers to the political context in which indicator information is read and applied. For those responsible for monitoring the state of their country's education system, the OECD indicators offer relevant information on a large number of aspects. This information is helpful but not sufficient as a basis for decision-making, however, and the set of indicators has to be supplemented with additional, country-specific data that make it possible to examine international indicators in relation to the particular concerns of various national audiences (United States, Special Study Panel, 1991).

Relationships among the indicators

Educational statistics appear in many forms – indeed, governments have collected data on education for a long time – but not all statistics qualify as indicators. As was noted previously in this chapter, the transformation of a statistic into an indicator is not primarily achieved through mathematical formalism or some numerical operation, but results from a complex interaction among cognitive and political processes.

The critical issue in the development of education indicators is not merely that of ensuring the validity of quantitative measurement in education. Despite the persistent scepticism among many educators about the use of algorithms and production functions as representations of complex educational processes, the feasibility of reliably measuring important aspects of such processes is by now widely accepted by research workers in the social and behavioural sciences. This does not mean, however, that all the epistemological and theoretical implications of this approach are well understood. Moreover, many technical problems still beset measurement in education. The analysis of qualitative variables in multilevel statistical models is only one example that comes to mind. Yet there has also been much progress, and this in turn has created increased awareness of the theoretical aspects that still evade clarification.

This leads to the proposition that the main difficulty of developing a set of education indicators derives not from technical issues but from the political context in which indicators are assigned a meaning. The development of an indicator stands or falls with the choice of the policy areas for which measures are needed. The inherently political nature of the exercise is exacerbated because an indicator is context-specific and, hence, needs to be interpreted as an element in a composite set of indicators that provides sufficient information on other aspects of the education system as well.

This is to say that an indicator alone is not informative. In order to qualify as an indicator, a measure should have an understandable relationship with other indicators. The criteria of choice therefore determine the coherence and usefulness of the "basket" of measures identified as indicators. Educational theory has not yet advanced to the point where heuristically valid, encompassing models can be proposed. Accordingly, it cannot yet be confirmed whether the indicators selected for inclusion in a set actually are the most relevant and policy-sensitive in their orientation. Many "productive" variables have been identified on the basis of theory and previous research, but it seems doubtful that these, collectively, can describe the "health" of the education system and provide a reasonably adequate understanding of its functioning. The conclusion must be that, although a single, multilevel model of the education system can of course be proposed, its adequacy will remain in doubt even if the model itself is subjected to an empirical test. It would therefore be highly pretentious to claim that the set of education indicators that has been constructed at the international level with the co-operation of the OECD countries is already perfect.

What is available at present is an embryonic rather than mature set of education indicators. Its improvement demands a lot of work: an international standard classification of education (ISCED) does exist, but it is obsolete and very difficult to use; educational statistics are abundant but only a fraction is appropriately used; some regularly collected statistics have lost their policy relevance, while no data are being collected for other policy areas that may be of interest. In short, the menu of data collection on education is outdated. These qualifications make it even more important that the realised indicators are presented coherently and in accordance with both theoretical principles and empirical knowledge.

If the indicators in the OECD set cannot be adequately interpreted in isolation from other indicators, then the assumptions concerning possible linkages should be made explicit. Moreover, these hypotheses should be tested using actual indicator data. This, in effect, contains a call for a research programme to improve our understanding and appreciation of the indicators and to offer guidance regarding their organisation and interpretation.

Under such a programme, it would be necessary to examine the theory that underlies the development of the OECD indicators, and to formulate hypotheses on how and why certain indicators may, or may not, be related. It would also call for the estimation of the interrelationships existing among the indicators in the set. Strategically, the objectives would be to implement procedures for quality control, establish criteria for the classification of the data, and to offer evidence useful in deciding on matters such as redundancy and the level of data aggregation. Ultimately, primary data analysis will be needed so as to determine the validity and accuracy of the indicators and, by drawing attention to their interrelationships, to facilitate their interpretation.

Conclusion

The risks and factors that caused the decline of the social indicators movement in the United States during the late 1960s and early 1970s clearly also pose a potential threat for the OECD education indicators. Fortunately, the participants and delegates attending the Lugano General Assembly of the INES project in September 1991, after a long, lively and, on some aspects very difficult debate, took the decision to publish a first set of international education indicators – despite the certainty that errors would occur, that many important areas would not be covered, and that it would be difficult to avoid misleading interpretation.

Both decision-makers and research workers recognise that our knowledge and understanding of the education system are partial and incomplete. Clearly, acknowledgement of the approximate nature of indicators that are based on a necessarily incomplete picture of education, and recognition that problems beset their interpretation, should not prevent the stakeholders from becoming accustomed to employing education indicators.

The concerted international effort to develop meaningful education indicators is not brought to a conclusion with the publication of the first set of indicators in *Education at a Glance*. On the contrary. With a tangible product in hand, and with the promise of "more to come", politicians and decision-makers in OECD countries are likely to put increased emphasis on the legitimate demand that the comparisons to be presented in future editions should be fair and correct, both methodologically and politically. Consequently, the first publication of OECD education indicators may well be seen as marking a new beginning.

BEST COPY AVAILABLE

Bibliography

BAUER, R.A. (1966), *Social Indicators,* Cambridge, MA: MIT Press.

BLOOM, B.S. (1976), *Human Characteristics and School Learning,* New York: McGraw Hill.

BURSTEIN, L., OAKES, J. and GUITON, G. (1992), "Education Indicators", *Encyclopedia of Educational Research* (6th edition), Vol. 2, pp. 409-418.

CARROLL, J.B. (1963), "A Model of School Learning", *Teachers College Record,* Vol. 64, pp. 723-733.

CARROLL, J.B. (1989), "The Carroll Model. A 25-year Retrospective and Prospective Review", *Educational Researcher,* Vol. 18, pp. 26-31.

FRASER, B.J., WALBERG, H.J., WELCH, W.W. and HATTIE, J.A. (1987), "Syntheses of Educational Productivity Research", *International Journal of Educational Research,* Vol. 11, pp. 145-252.

HAERTEL, E. (1988), "A Technical Review of the National Assessment of Educational Progress: Introduction and Overview", *International Journal of Educational Research,* Vol. 12(7), pp. 673-678.

HAERTEL, G.D., WALBERG, H.J. and WEINSTEIN, T. (1983), "Psychological Models of Educational Performance: A Theoretical Synthesis of Constructs", *Review of Educational Research,* Vol. 53, pp. 75-91.

HANUSHEK, E.A. (1979), "Conceptual and Empirical Issues in the Estimation of Educational Production Functions", *Journal of Human Resources,* Vol. 41, pp. 351-388.

HARNISHFEGER, A. and WILEY, D.E. (1976), "The Teaching-learning Process in Elementary Schools: A Synoptic View", *Curriculum Inquiry,* Vol. 6, pp. 5-43.

HUSÉN, T. (1967), *International Study of Achievement in Mathematics: Comparison of Twelve Countries,* Stockholm: Almqvist and Wiksell; New York: Wiley.

KIRST, M. (1989), "Accountability: Implications for State and Local Policymakers". Paper prepared for the Office of Educational Research and Improvement, December 1989, Washington, D.C.: United States Department of Education.

OECD (1973), "A Framework for Educational Indicators to Guide Government Decisions", Paris: OECD (document for general distribution).

OECD (1992*a*), *The OECD International Education Indicators: A Framework for Analysis,* Paris: OECD/CERI.

OECD (1992*b*), *High-Quality Education and Training for All, Part 2.* Paris: OECD.

ROCKWELL, R.C. (1989), "Lessons From the History of the Social Indicators Movement". Paper prepared for the Special Study Panel on Education Indicators, Washington, D.C.: National Center for Educational Statistics, U.S. Department of Education.

UNITED STATES NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION (1983), *A Nation at Risk: The Imperative for Educational Reform,* Washington, D.C.: Government Printing Office.

UNITED STATES SPECIAL STUDY PANEL ON EDUCATION INDICATORS (1991), *Education Counts: An Indicator System to Monitor the Nation's Educational Health*, Washington, D.C.: National Center for Educational Statistics, United States Department of Education.

van HERPEN (1992) "Conceptual Models in Use for Education Indicators", in *The OECD International Education Indicators: A Framework for Analysis*, Paris: OECD, pp. 25-51.

Chapter 2

Observations on the Structure, Interpretation and Use of Education Indicator Systems

by

Anthony Bryk** and **Kim Hermanson
University of Chicago, United States

New interest in education indicators

The second half of the 1980s saw a new interest in education indicators, as witnessed by an increasing number of publications. Organisations such as the U.S. Department of Education, the National Science Foundation, the Council of Chief State School Officers (CCSSO), the National Research Council of the National Academy of Sciences, the Rand Corporation and nearly all state governments are currently involved in developing and improving education indicators. "Hardly an educational group or agency at the national or state level has not become involved in the business of education indicators during the 1980s" (Smith, 1988, p. 487).

In the United States this surge of interest can be traced at least back to a 1983 report, *A Nation at Risk*, which triggered broad public concern about education and initiated a strong push for closer monitoring of the system, its schools and personnel. In 1984, the Secretary of Education's Wall Chart compared states' educational performance. The Wall Chart prompted the CCSSO to begin work on a fairer set of indicators and to create the State Educational Assessment Center. In the following year the National Research Council recommended that data collection and reporting be reorganised under a strong federal agency.

International efforts to create education indicators received a new impetus in 1987 when the United States government supported a cross-national indicator project established under the responsibility of the OECD, the International Education Indicators Project. At the same time, the Hawkins-Stafford Act of 1988 provided congressional support for national indicator development, national and state co-operation in data collection, and the expansion of the National Assessment of Educational Progress to include state-by-state comparisons. In 1989, the National Forum of Education Statistics was created to recommend improvements to national statistics, and a congressionally mandated panel of experts and policy-makers was formed to report on data quality and recommend further improvements (Shavelson, 1987; Burstein *et al.*, 1992). Meanwhile, the increasing array of data and their importance to policy-making have raised concerns about quality, relevance and interpretability.

Amidst this burgeoning governmental activity, there has been growing discussion of the potential of indicator systems to improve schooling. Kaagen and Coley (1989) describe indicators as diagnostic tools that would offer "a unique opportunity for state policy-makers to affect local education practice in a most efficient way" (p. 12), and they affirm that the main purpose of these systems is to "assess direction, mission, and strategy" (p. 9). The CCSSO argued that "state-level policies must be driven by the need to improve the education system as suggested by these outcome measures" (1989, p. 5). More generally, indicators are promoted as efficacious instruments with which to monitor the educational system, evaluate its programmes, diagnose its troubles, present solutions for reform, and hold school personnel accountable for the results – a truly impressive array of tasks.

This chapter addresses issues that arise when interpreting and analysing information produced by a system of education indicators. The focus is on some very basic questions: Analysis and interpretation for what purpose? Rhetoric aside, how will this information actually be used? Can it catalyse school improvement in the ways suggested? These questions lead to reflection on some fundamental premises concerning the nature of schools as organisations; about the exercise of control over these institutions and their processes; and about how information can productively enter this domain.

These topics are briefly considered below. Attention is also given to contemporary discussion of indicator systems, which appears to encourage a rather simplistic view of schools, how they are controlled, and how new information might influence future activity. More specifically, this discussion tends to assume that school operations can be adequately represented through a production function model. In this model, indicators would tap into critical processes, and the resulting information would be used directly by external policy-makers to control schools through instruments such as rules, administrative sanctions, and incentives. This direct use of indicator data is referred to as the "instrumental use" model, and it recalls the "experimenting society" (Campbell, 1969; Rivlin, 1971) whose promises of advancing social betterment by using the modern tools of programme planning, budgeting, evaluation and analysis remained unfulfilled. However, this production function model of the school and the instrumental model for indicator information are perhaps most usefully viewed as a foil.

Finally, an alternative conception, which may offer a more prudent set of aspirations for an education indicator system, is sketched out. It relies on a view of the school as a social system where personal interactions are primary, where structural reform often requires changing the values and tacit assumptions that underlie these interactions, and where the primary purpose of new information is to foster an informed, sustained discourse about the means and ends of education.

Much of this chapter is cautionary in tone. It articulates some worries and some caveats and reflects some uncertainty about this new policy initiative. To be sure, education indicators hold considerable potential for advancing education, but they also offer much opportunity for misuse.

Basic premises concerning schools and schooling

Schools as organisations

Bidwell (1965) described two different ways in which schools have been conceptualised in educational research, policy and practice. Over the last two decades, the rational/

bureaucratic model has clearly dominated. In this view, the school is a formal organisation characterised by: a functional division of adult labour into specialised tasks; teaching roles defined by subject matter and type of student; emphasis on social interactions that are rule-governed, affectively neutral and with limited individual discretion; and a form of authority that is attached to the role within the organisation rather than to the person. The bureaucracy is managed by a specialised administrative staff that exercises control through a variety of instruments: job definition; rule formulation; and communication and information networks, including accountability systems. Most current discussion of education indicators occurs in this area, where the data are seen as the newest management tools for the policy analyst and administrator.

Although much can be said in support of the idea of the school as a formal organisation, it has unfortunately tended to undervalue certain elements that are central to the second conception – that of the school as a "small society." This model favours a more diffuse adult role and a minimal division of labour. It emphasizes the informal and enduring social relationships on which teaching and learning draw and the influence of the school's normative environment. It also focuses attention on the tacit beliefs that are conveyed through personal interactions.

While deference to formal authority is often the explicit mechanism of control in these "small societies", this authority rests on a set of shared views, and the behaviour of school participants is largely autonomous. If indicators are to be of any value to this conception, they would most likely serve to change the shared values and assumptions that are the basis of much of the day-to-day life in schools.

Multiple, diverse and sometimes conflicting goals

Discussion of education indicators typically assumes that there is some generic academic achievement outcome (or a small set of outcomes) which good schools seek to maximise. However, the nature of authentic academic achievement is actively debated (Gardner, 1983; Resnick, 1987; Archbald and Newmann, 1988). While few would claim that standardized tests are of no value in assessing knowledge and academic skills, reformers tend to emphasize the need to refocus instruction on immersion in specific subject matter, on forming students to be active learners, and on engagement in substantive conversations to deepen understanding and promote communication skills. Clearly, such a conception of student achievement has vast and largely unknown implications for any indicator system that seeks to take these aims seriously.

In addition to academic achievement, schools have other objectives. They seek to promote a wide range of social competencies, from taking turns to working co-operatively in small group settings. They are also responsible for developing caring human beings and forming an engaged public citizenry, goals that are fundamental to civic life. Beyond this, schools are increasingly asked to redress a vast array of social problems, from driver education to drug abuse and teenage pregnancy. These varied demands must be sorted out and incorporated into the functioning of each school.

Although discussions of education indicators may acknowledge the diverse aims of education, it is nonetheless commonly presumed that it is possible to focus on the "core of schooling" – academic achievement and the processes instrumentally linked to it – while ignoring everything else. Such thinking implies a segmentation in the organisation, operation and effect of schools. However, there is growing evidence, for example, that the social structure of schools influences student engagement and teacher commitment, both of which are linked to students' academic achievement (Bryk *et al.*, 1990). Further, it seems likely that what is involved is more than reciprocity of outcomes, where a positive result in one area begets improvements in another. Rather, it may be that the social structures which advance the diverse aims of schooling are themselves interrelated.

The systemic nature of schooling processes

It follows that it is important to consider the conceptions of schooling processes that underlie the current discussion of education indicators. They have as their basis the mechanistic production function model that is the dominant paradigm in quantitative social science. They assume that the operations of schools can be viewed as an ensemble of linear, additive chains of cause and effect, where deficiencies can often be identified with a specific single factor or small set of factors. It is argued that indicators can help identify these weaknesses and allow policy actors to overcome them.

In contrast, a systems theorist might describe schools in more organic terms (Sutherland, 1973), as flows of information and communication involving a variety of feedback loops. "When a system breaks down, the breakdown is usually caused by multiple factors that may amplify each other through interdependent feedback loops" (Capra, 1982, p. 269). Interaction is fundamental to the operation of such systems, so that even when an action has certain intended effects, it will often also have a wide range of repercussions throughout the system, some of which may be very undesirable.

Clearly, this effect raises questions about how indicator information, developed around certain aims and activities of schools, can be used to influence overall system operations positively. The strong, direct and independent causal connections assumed in the production function model underlie the expectation that policy action based on indicators will produce the intended consequences. In contrast, if schools are more like organic systems, then such policy actions may result in much more complex and indirect consequences.

Proceeding one step further, schools may also be legitimately described as social systems in which technology plays only a modest role and in which people change people through personal interaction. Effectiveness depends largely on the efforts of social participants and their mutual trust. Stated somewhat differently, schools are places where personal meaning and intentions matter, and this forms the basis of their effects (*cf.* Bryk and Driscoll, 1988; Bryk *et al.*, 1990).

Although structural-functional models of schooling can afford powerful insights, they often fall short of explaining human action. For example, organisational size is strongly correlated with the bureaucratisation of relationships and the attendant consequences of impersonal treatment, detachment and alienation. Yet there are large schools where this does not occur, and small schools where it does. Strongly held beliefs (such as "we want this to be a place where people know and care about each other"), and the actions deriving from them, can markedly affect otherwise rigid structures. A large school size may work against personal treatment, but determined human action can overcome this.

The central point is that social systems operate according to certain basic ideas about what is proper, right and just. For example, the structure of American education is largely designed to offer opportunities for individual self-expression, evidence for which can be found in the broad and diverse high school curricula, in students' right to exercise choice in their study programmes and teachers' demands to control subject matter and how it is taught. This educational premise has its source in the basic American cultural value of expressive individualism (Bellah *et al.*, 1985), which influences both the nature of students' engagement in schooling and teachers' willingness to expend effort on their work. However, different contexts are likely to produce different patterns.

This last observation has important implications for an international indicator system. To obtain useful cross-national comparisons, a common data collection and reporting framework, based on some assumed model of schooling, must be established. Great care is required, however, when interpreting and using this information, as it can easily lead to facile generalisation where more subtle interpretations are required. Prudent use of indicators requires that the information be recontextualised so that differences among countries can be understood. The danger is that the indicator model will frame the discussion, in essence becoming the assumed model for schooling in every country. In that case, the common conceptual structure of the indicator system will have an undesirable effect on action.

For example, facile comparisons are made of the American and Japanese school systems. It is argued that the Japanese "have larger class sizes and still do better" and that "their teachers and students work harder". While this may be true, there are also distinctive cultural factors at work. Rohlen's study of Japan's high schools (1983) is instructive in this regard, as it documents profound differences between the two countries with respect to individual and collective values, individual rights and social responsibilities. To make the American system more like the Japanese would require more than just manipulating the conventional instruments of educational policy. It would also entail direct and sustained attention to these value structures.

Control over schooling

The development of an indicator system also requires careful consideration of how the organisation and processes of schooling are controlled. The indicator movement argues that these data will increase accountability and lead to new policies to improve education. While the principle is broadly acclaimed, the precise mechanisms are left largely unaddressed. For this reason, attention should be given to the basic mechanisms at work in controlling education.

Weiss (1989) argues that five diverse mechanisms interact to control schooling. The first involves the political control exerted by elected officials and thus indirectly by citizens. Such control operates through budgeting, policy-making, and the formulation of procedural regulations; it focuses on the accountability of schools to the public. Such political control is an external effort to direct the internal operations of schools towards objectives held important, in ways deemed appropriate.

In contrast, the bureaucratic hierarchy – staff of school systems and departments of education and education offices in central governments – exercises administrative control through procedures such as hiring of personnel, performance evaluation, rule interpretation, and information management. It is assumed that such managerial control over employees and resources directly influences what occurs in schools and the results they produce. Administrative control is exercised at the individual level (as managers construct incentive and control systems to induce effective work), at the organisational level (through job descriptions, a formal division of labour, established work routines, and the like) and at the inter-organisational level (through exchanges of influence, resources and information among interrelated enterprises).

The third form of control evolves from the observation that schools, even American public schools, are subject to a number of market forces. Schools, for example, compete for staffing with other school systems, for public resources among themselves and with other social welfare initiatives and to some extent, through location, for students. In contexts where explicit choice is possible, such as magnet programmes and private schools, market influences are even more substantial. Such schools must ensure that sufficient numbers are willing to enrol and thereby generate the necessary resources if they are to continue to exist.

The fourth mechanism focuses on the control, shaped by training and norms of practice, which professionals exert over their own work. It operates through specialised pre-service, continuing education and a wide range of professional associations. Because teachers and administrators have special expertise, they expect freedom to exercise judgement over professional matters. Much classroom decision-making, for example, is not directly affected by other control mechanisms. Similarly, teachers and administrators tend to resist outside control by non-professionals, who are considered unqualified to judge.

Finally, values and ideas also exercise control over schools. This mechanism is closely connected to the notion of schools as systems in which embedded normative relationships play a substantial role. Control can be exercised over schools by encouraging people to "come to think differently about their situation – believing that more desirable means exist to achieve ends, or coming to value different ends" (Weiss, 1989, p. 20). Arguably, this form of control is both powerful and illusory. It is difficult to imagine sustainable social change that is not preceded by a reorientation of values and a commitment to new ideas, and such change is likely to occur slowly and irregularly.

An uncertain connection: indicators and educational improvement

Some advocates of education indicators, such as Kaagan and Coley (1989), the CCSSO (1989, 1990) and Smith (1988), envision instrumental use of the data, primarily through political and administrative control mechanisms. In this view, an indicator system is structured around a model of schooling that captures the nature of academic outcomes and simultaneously monitors key inputs and processes. Moreover, the relations among inputs, processes and outcomes are empirically substantiated, and this justifies policy formation, rule writing, or use of incentives or administrative sanctions that can be directly linked to the specific processes needing improvement.

The technical requirement that all components of the indicator system be based on established means-ends relations has important implications, because it can limit the information that enters the policy arena. To the extent that indicator information has greater credibility than less formalised knowledge, such as case studies and clinical expertise, it can be a potent source of conceptual distortion in subsequent policy formulation. Of concern here are basic issues about the nature of social science evidence and about how indicator information can advance educational policy and practice.

Limitations imposed by social science

"What we know", in a rigorously scientific sense, is limited by existing research techniques. The linear, additive, unidirectional models that are the stock in trade of the quantitative social scientist are too simplistic for the phenomena studied. No serious analyst, if pressed, is likely to maintain that social reality is a simple assembly of non-interacting additive components. Constraints, tensions, dilemmas and conflicts are routine elements of discourse about schools, yet they are virtually absent from empirical models.

Substantive limitations are also imposed by the unexamined values held by social scientists. Allegiance to a particular theory, such as rational choice, can cause distortions of perception; some phenomena might be more properly viewed in a different light. As a result, whole domains of school life may be omitted from inquiry. For example, the formation of personal traits, such as a sense of craft, are important aims of education (Green, 1985). Yet these concepts receive scant empirical attention. Much the same can be said about aims such as promoting social responsibility and civic participation. Similarly, whole domains of schooling, such as extra-curricular activities, are largely ignored.

In other instances, the findings of some well-established research traditions run so counter to common sense that many observers conclude that they must be wrong. Studies showing a lack of relationship between fiscal resources and student outcomes are an example (see Hanushek, 1981, 1986). While in a narrow sense the findings are correct – aggregate resources bear only a weak relation at best to aggregate academic outcomes – the real issue is proper conceptualisation (Bidwell and Kasarda, 1975). The school and district-level models typically used in this type of research hide enormous variation in the ways resources are allocated within these units, which is what actually affects individual students' opportunity to learn and achieve. This kind of information is necessary in order to discern how to allocate and use fiscal resources more effectively.

At best, a well-designed indicator system would give a relatively coherent representation of a segment of the educational system and would offer solid scientific evidence that the inputs and processes studied are linked to the measured outcomes. The model, however, would be far from complete (MacRae, 1988). To be complete, it would have to have an explicit representation of the diverse aims of schooling, the means-ends linkages for each of these aims, and the interrelationships (including tensions, conflicts, and dilemmas) that might exist among these multiple, interwoven educational processes.

Such a model would require a multi-level formulation, including at least: classroom-level concepts about student learning, teaching pedagogy and classroom practice; school-level concepts about curriculum organisation, academic and disciplinary policy, quality of social relations, adequacy of available resources and school leadership; and key concepts that capture the major support and administrative functions at the district and higher levels of government. In addition, the component focusing on student experiences and outcomes would need an explicit developmental dimension. Interrelationships among experiences in the first five years of life, learning in elementary and secondary schools, and adult outcomes including active citizenship, productivity in the work place, and personal well-being, would also have to be specified.

Social science has helped to establish many of the important components to be included in such a model, but it has identified neither the full range of components nor their interaction in an operating educational system. In short, social science knowledge of schooling is both partial and fragmented. It is far from an integrated theory of school organisation, processes and effects. Consequently, any indicator system based on this knowledge will be incomplete. While it certainly seems worthwhile to strive to construct more complex models of the educational system, it is a delusion to presume that a comprehensive model of sufficient complexity to support instrumental use can be developed soon.

The foregoing analysis suggests that attention should be focused on how limited indicator information can be recontextualised so as to allow for a broad understanding of schools and the schooling process. Educational philosophy, school case studies and clinical expertise are all needed to illuminate the indicator data, and more generally, to enrich the debate on school improvement. To be sure, many of the ideas introduced into the discussion will lack scientific warrant. As a result, issues of judgement, expertise and, in some instances, taste, will naturally arise. Since such matters are ill-suited to the bureaucratic control of schooling envisioned in the instrumental model, attention must also focus on the role of information in shaping future educational practice.

The limits of external control through indicator data

The instrumental model assumes that the actions of managers and policy-makers, based on an understanding of a segment of the educational system, are benign with respect to all other segments. This assumption is contradicted by much accumulated evidence about government efforts to manipulate schools. While instances of successful policy action can be identified, it is also clear that many initiatives have had extensive unintended consequences.

When schools are analysed as social systems, the problem is amplified. Policy changes often introduce new structures that must be adopted if the policy initiative is to be effectively implemented. For example, the move toward greater school autonomy carries the demand that teachers accept responsibility for the academic development of every child. Instead of defining acceptable performance as the meeting of stated standards, teachers should become outcome-oriented. However, the processes necessary to promote such a change in ideas are seldom explicitly provided for in policy formulation. As a result, instead of promoting values consistent with the new policy intent, schools often respond marginally to the new initiative and introduce other changes that may lead to undesirable outcomes.

High stakes testing programmes are a case in point. While intended to improve school accountability and, through this, student learning, they may also alter schooling in undesirable ways. Critics fear that teachers will begin "teaching to the test" which, they argue, is not appropriate to the aims of education. Others worry that increased accountability will cause schools to change their policies concerning school admission and expulsion (*e.g.* encouraging students who might do poorly to go elsewhere). Worse yet, schools may become more calculating in how they focus attention, perhaps ignoring students who are likely to leave since they "won't count" in the final report. In short, unless teachers endorse accountability as a legitimate dimension of their work, such external policy initiatives may lead to frustration.

Implications for indicator systems

Prudent aspirations

In the late 1960s, many thought that chronic social problems could be solved through the efforts of social research. Social science would provide the link that would make it possible to solve social problems, just as engineering provides the link between the physical sciences and a wide range of material problems. Decision-makers would identify a specific social problem requiring a solution, and researchers would be called upon to identify the lack of understanding that generated the problem and provide the missing knowledge or to assess the relative effectiveness of a number of possible solutions. It was assumed that once this knowledge was furnished, the appropriate policy response would follow directly (Riecken and Boruch, 1974; Albaek, 1989).

These aspirations were naive on several accounts: they failed to comprehend the limited and often uncertain character of much social science knowledge; they ignored the influence of political processes and interest groups on social problem-solving (by invoking the individual decision-maker); and they were virtually silent on issues of organisational control and on how the beliefs and values of administrators shape policy implementation (Lipsky, 1980). Much of the current debate about education indicators appears to recreate this same "engineering model" of the relationship of information to action. In reflecting on this flurry of activity, Cronbach (1982) called for more prudent use of expert knowledge in decision-making. His comments seem equally appropriate today:

> "Findings of social science can rarely or never identify 'right' courses of action ... Our main stock in trade is not prescriptions or laws or definitive assessments of proposed actions; we supply concepts, and these alter perceptions. Fresh perceptions suggest new paths for action and alter the criteria for assessment ... Quasi-prescriptive conclusions, on the other hand, invite disbelief." (p. 72)

An enlightenment function

These comments build on Weiss's argument (1977) that the relationship of social science to social problem-solving is more akin to "enlightenment" than to "engineering". Education indicators are of value because they can increase the understanding of problems and introduce new ideas. They can signal new problem areas, offer conceptual frames, provide useful information for brainstorming about possible solutions, and, more generally, inform the broader public (Lindblom and Cohen, 1979; Saxe, 1986; Andrews *et al.*, 1990).

According to this view, information from indicators will rarely provide specific solutions to problems of school improvement that can be directly enacted. Rather, indicators are useful for "pre-policy" formulation. The data give information on results, provide information that can help define problem areas meriting closer attention, and stimulate discussion about possible solutions (de Neufville, 1975, p. 44).

BEST COPY AVAILABLE

A specific objective: enriching public discussion

Indicators are not the exclusive domain of policy-makers and administrators who seek to increase external control over schools. They also offer a means to educate the community, as they can encourage public involvement in educational issues. The issue, in this regard, is not whether indicators are discussed by the media, but whether the information they produce stimulates discussion and deepens understanding of the aims and means of education.

Much of the current promotion of indicators has an explicitly political purpose, which is to sustain public debate on the issues and problems of education in order to maintain education's place on the public agenda and its claim to resources. As such, it is an attempt to break the established cycle in education of heightened short-term concern followed by long periods of neglect.

Indicators must be designed so that they truly inform. Composite indicators, and especially the "mega-indices" aspiring to become the Dow Jones of education (Guthrie, 1990), may not meet this criterion. They are likely to attract media coverage and may galvanise public attention, and this is important. However, they appear to assume that "a few simple numbers are needed to fire the passions of the masses" because that is all they need or can digest (Cremin, 1988, p. 255). In this "progressivist" vision of a democratic society, technical expertise is considered central to social betterment, but if the indicators involved are not detailed they cannot educate. Knowledge has to be broadly held, if democracy is not to become a technical aristocracy. Over 70 years ago, James Harvey Robinson (1923) noted that if formal knowledge is to have social value in a democratic society, it must be humanised:

> "The results [of attempts to humanise scientific knowledge]... seem very inadequate compared with the possibilities ... It has become apparent that we must fundamentally reorder and readjust our knowledge before we can hope to get it into the current of our daily thought and conduct. It must be re-synthesized and re-humanised. It must be made to seem vitally relevant to our lives and deeper interests." (p. 100)

Seeing indicators as promoting informed public discussion claims a constructive role for them, while acknowledging their limitations. It is a prudent aspiration, consistent with past experience about the role that social science can play. Moreover, it is a proper aspiration, in that it respects the role that knowledge about education should play in a modern, democratic society: not to manipulate the public or empower a technical elite, but rather to enrich sustained public discussion of schools and the processes of schooling in order to improve them.

In this perspective, indicators primarily influence education not through direct administrative action but by contributing to changes in ideas and values. Although the process may be lengthy, the consequences can be quite powerful. When discussion is influenced by the reporting of indicators, virtually every other control mechanism will eventually be affected. Policy-making activity may be refocused, as administrators start to redefine problems and reorient actions. However, such developments are neither quick nor by any means assured.

Structuring the system: an information pyramid

This view also implies a need for an appropriate reporting system for education indicators, so that diverse audiences are engaged and truly informed. Several observations seem pertinent here. First, such a system is likely to require a number of different reporting levels, so as to facilitate use by audiences with different interests and expertise. Second, the levels should be interrelated, so that when a trend in a particular indicator draws attention, it should be possible to move easily to detailed statistical information, including in-depth studies and inspectorate reports, that might illuminate some of the forces at work. Third, the system should have a strong conceptual organisation, that captures both established means-ends generalisations from social science and the best clinical expertise. Although a single comprehensive model of the education system and its processes does not exist, the reporting system could be built around an interrelated set of problems and educational concerns.

Each report might then resemble an information "pyramid". At its top would be a limited number of key indicators of status (and when presented as a time series, of progress) in some domain. For example, if a major reporting area is students' readiness for school, three or four key indices might summarise students' cognitive, social, health and nutritional status at entry. Each might be composites of more detailed statistics. They would show the current situation at the aggregate level and thus focus attention on concerns that might merit closer scrutiny. This level of information is for the broadest public; it is the kind of data that might appear in a graph in newspapers or be briefly mentioned on the evening news.

The next two levels of the pyramid elaborate on these summary indicators. At level two, a carefully chosen expanded set of statistics would afford a more in-depth understanding of the forces at work behind the key indicators. This information would be a resource for brainstorming about future policy efforts – what is described above as "pre-policy" formulation. On the "readiness for school" topic, this additional information would have a developmental perspective. It might include statistics on children with low birth weights, children born with drug dependency, children living in poverty and various at-risk family structures. Data on access and quality of pre-school and basic health care would also be included.

Level three would add a further dimension by reporting on selected research studies, including case studies, programme evaluations, and small-scale quantitative studies. Craft knowledge, as captured in inspectorate reports, could also be included. Although the scientific warrant attached to level three would necessarily be somewhat weaker than that of the other two levels, this information may be very useful in suggesting ideas and solutions to problems.

Interpreting and reporting statistical results

Technical issues also need to be considered. These include standards of comparison for interpretation and decisions about the specific types of statistics to be reported.

Standards of comparison

Formally, an education indicator is a statistic covering a set of units that conveys information about a condition in the education system. Evaluating this information, however, requires reference to some standard of comparison. There are several obvious possibilities: a comparison to self over time (the national development model); comparison with other countries (the "Olympics" model); and comparison against some externally defined standard (the educational goals model). Depending on the standard of comparison, different technical issues arise.

Comparisons over time place a premium on the stability of the indicator system in terms of data definitions, measurement processes, sampling frames, and the organisational methods of data gathering. Corruption of the processes over time is always a potential problem, particularly if the indicator data are used for accountability purposes. The integrity of the indicator system depends critically on indicators being just that – indicators – and not information used for micro-management.

Comparisons with other countries raise a somewhat different set of issues. As noted earlier, they require a common data collection and reporting framework based on some assumed model of schooling. While such data can help to identify and clarify differences for the dimensions measured, the information must be interpreted with great care because of differences of structure, language and culture.

Care must be taken in building models of this type, because whenever a set of variables is collected, there is a natural temptation to infer causal connections. For example, when the U.S. Secretary of Education's first Wall Chart was released, some were quick to point out that "educational resources didn't matter" because the state of New Hampshire had the highest SAT scores and very low per pupil expenditures. What the "armchair analyst" failed to recognise is that the sample of students taking the SAT in New Hampshire is relatively small and highly biased by the large number of elite private schools in the state.

The potential mischief here is enormous, whether one engages in simple comparisons, as above, or in sophisticated statistical modelling with structural equation techniques. Quite simply, indicator data cannot support causal analysis. The data are aggregate statistics about countries, while the inferences are typically about educational processes affecting students. This difference in levels between where the indicators are reported (countries) and where the inferences are to apply (students and schools) results in a classic logical fallacy (*cf.* Burstein, 1980).

In terms of external standards of comparison, such as national educational goals, the main issue is that different countries are likely to formulate goals differently and impose different criteria for judging success. Further, these goals are likely to change over time. However, to the extent that stability in the indicator system is essential for comparisons over time, appropriate modifications may be difficult to achieve.

Types of statistics reported

Typically, indicators are based on simple descriptive statistics of central tendencies such as means and ratios. When reported on a country-by-country basis, they provide information about the relative standing of countries on a particular dimension. In other cases, indicators are adjusted statistics designed to control for one or more factors. For example, a measure of "national commitment to education" could use a ratio of per pupil expenditure relative to per capita income, or gross educational expenditures relative to gross national product, or even starting salaries for teachers relative to those in other selected professions. The idea is that the indicator should give information about each country's valuation of education relative to other activities.

Adjusted statistics are also used when indicators are to report on the value added by schooling (*i.e.* by taking account of personal background and prior experience) or when "system efficiency" statistics (*e.g.* the amount of output per resource unit) are desired. In this regard, the OECD might consider developing an indicator of student gains in achievement over time. Unlike status measures, which may be based on very different configurations of students at different grade levels, reporting of gains provides direct evidence of the relative productivity of different levels of the educational systems.

Also worthy of consideration are measures of diversity. For example, in addition to reporting mean levels on student outcomes, the spread in achievement could be measured. When viewed as a time series, the data would make it possible to examine whether certain national systems are becoming more or less diversified over time. Cross-national comparisons would make it possible to examine whether greater internal diversity accompanies high mean levels of outcomes. Are there some national systems that achieve both high mean levels and have minimal internal diversity? Although these questions are framed in the context of student outcomes, similar questions can be asked about school structures, resources and instructional processes.

Going one step further, indicators might also address equity considerations, for example with regard to gender and income level, school size and location. Both time series and cross-national comparisons could be instructive here as well.

The indicator system in the broader social framework

Nurturing the technical and information user communities

Creating an indicator system also requires considering the social infrastructure that surrounds it. Clearly, attaining substantive richness will require strong ties between those charged with maintaining the indicator system and the research community. The latter group should provide the concepts and instruments used at levels one and two and conduct the research that constitutes the interpretative framework at level three. Further, if the indicator information is to catalyse broad public discussion, then it must also attend to the mechanisms that initiate such discussion. Likely stakeholder groups must be identified and included in the process of developing the system, so that they will have an interest in it. The individuals and organisations that represent "craft knowledge" are particularly important in this regard (see Cohen and Spillane, Chapter 16 of this volume).

The indicator agency and the public trust

The agency that develops and maintains the indicator system is entrusted with an important public responsibility. System design is not merely a technical act; it is also a fundamentally political one. The content of the system will shape future discussions about education because it will help identify the problems that merit attention and the solutions that are advanced as warranting support. The indicator system can have a significant effect on public life, albeit through the rather subtle mechanism of shaping ideas and values.

This suggests that some attention should also be paid to the organisation charged with this activity and the standards maintained by its personnel. Four principles would appear appropriate: pluralistic debate and inquiry; technical competence in design; independence from short-term partisan political interests; a prudent view of the role of information in education.

Schools are distinctive social institutions placed between the family and the community and serving both public and private goals. By virtue of their location and their mission, their aims and methods are likely to be continually debated by a number of concerned constituencies. Since the information system will shape the nature of these debates, it must devote attention to fairness towards all relevant interests. Without care and sensitivity, seemingly technical decisions can advantage some interests and disadvantage others.

The development of an indicator system will draw upon an array of substantive and methodological expertise. This expertise must be used not only for forming the system but also for sustaining it. There is a disturbing tendency in American education to separate the applied activities carried out by contract research firms from the more basic research conducted largely in universities. The ties between the indicator agency and the research communities on which it depends must be strengthened, so that the indicator agency will not be deprived of access to needed expertise.

Although indicator data will surely be used for partisan political purposes, the long-term health of the system requires that the agency involved be assured independence from short-term political influence. A non-partisan mission is essential to the integrity of this system. Governments shift from time to time, but the public will be well served only if integrity is maintained across these changing influences.

The primary commitment of the agency must be to the long-term development of the system of indicators and the improvement of education through good information. There is always a great danger that this aim will be subverted by short-term interests. The primary concern must always be the indicator system's constructive long-term contribution to the improvement of education. After all, no public institution requires more sustained collective effort than the schools. They are the bedrock of a free, vital, democratic society.

Conclusion

This chapter has argued that the analysis and interpretation of indicators is shaped by the structure of the indicator system. Key issues involve assumptions about the nature of schools and the processes of schooling, the aims of education, the exercise of control, and the underlying social relationships in the domains of policy, practice, outside expertise and the wider public.

While the discussion is based on more than two decades of American experience in using social science as a vehicle for educational problem-solving, it also reflects on a larger set of considerations about how schools are governed. As Cohen and Spillane (Chapter 16) note, the American system is relatively decentralised, and policies typically emerge through competition among diverse interests. Although the "genius" of the American system may be compromise, the resultant policy is not necessarily intellectually coherent. The system has a preference for objective "scientific data" over craft knowledge, and it lacks a strong and independent civil service in education administration.

Clearly, there is significant variation among the OECD Member countries in these areas. As a result, some of what is said above, especially on matters of control, will warrant consideration in light of each country's governmental structures and traditions. On the other hand, observations about schools as organisations, the schooling process, the nature and limits of social science knowledge and its association with craft knowledge to solve problems seem more universal.

In conclusion, it may be noted that indicator systems of the type described above will take time to develop. Not surprisingly, there are few quick solutions in education, and a good indicator system is surely not one of them. Past experience with indicators in other fields substantiates this point. As Murnane (1987) comments:

> "[There is a] striking dynamic interaction between the indicators available at any time, and analyses based on these indicators. Improved indicators lead to new analyses that raise new questions and call into question the usefulness of the new indicators. But the new questions would not have been identified without the improved indicators... Users of indicators have come to appreciate the importance of things not captured by existing indicators. In other words, sophistication leads to dissatisfaction, which leads to re-defined goals. These evolving definitions of goals have created new demands for indicators." (p. 102)

By its nature, an indicator system demands parsimony. It can provide useful information about the current situation, but moving from this limited knowledge to what should be done instead is a complex matter requiring much judgement and understanding. "More" information is not always "better". The ultimate long-term test of the indicator system to be developed by the OECD is not only whether it helps the audience to be better informed, but also whether it helps it to act more prudently.

In the short term, the best intermediate test may be found in the answer to the question, "Is public discussion enriched (or impoverished) by this new information?" Here it is worth recalling important human qualities which cannot adequately be measured, such as resilience to stress, courage in facing uncertainty, and caring and humanity in interactions with others. The great danger presented by indicators is that they narrow public discussion of the purposes and means of education to what can be measured, while ignoring those invaluable aspects that cannot.

Acknowledgements

In drafting this chapter, we have drawn heavily on discussion with members of the Special Study Panel on Education Indicators for the U.S. Department of Education. Extensive conversations with Duncan MacRae have been especially informative. We have also drawn liberally on ideas offered by Leigh Burstein, Dennis Jones, Art Powell and Floraline Stevens.

Bibliography

ALBAEK, E. (1989), ''Policy Evaluation: Design and Utilization'', *Knowledge in Society*, Vol. 2 (4), pp. 6-19.

ANDREWS, F.M., BULMER, M., FERRISS, A.L., GERSHUNY, J., GLATZER, W., NOLL, H., INNES, J.E., JOHNSTON, D.F., MACRAE, D., VOGEL, J. and WARD, M. (1990), ''Whatever Happened to Social Indicators? A Symposium'', *Journal of Public Policy*, Vol. 9 (4), pp. 399-450.

ARCHBALD, D. and NEWMANN, F.M. (1988), *Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School*, National Association of Secondary School Principals, Reston, VA.

BELLAH, R., MADSEN, R., SULLIVAN, W.M., SWIDLER, A. and TIPTON, S.M. (1985), *Habits of the Heart: Individualism and Commitment in American Life*, Harper and Row, New York.

BIDWELL, C.E. (1965), ''The School as a Formal Organization'', in March, J.G. (ed.), *Handbook of Organizations*, Rand McNally, Chicago, IL, pp. 972-1022.

BIDWELL, C.E. and KASARDA, J.D. (1975), ''School District Organization and Student Achievement'', *American Sociological Review*, Vol. 40, pp. 55-70.

BRYK, A.S. and DRISCOLL, M. (1988), *The High School as Community: Contextual Influences, and Consequences for Students and Teachers*, National Center on Effective Secondary Schools, Madison, WI.

BRYK, A.S., LEE, V.E. and SMITH, J.B. (1990), ''High School Organization and its Effects on Teachers and Students: An Interpretative Summary of the Research'', paper prepared for the Invitational Conference on ''Choice and Control in American Education'' at the Robert M. LaFollette Institute of Public Affairs at the University of Wisconsin-Madison, 17-19 May.

BURSTEIN, L. (1980), Multi-level Analysis in Educational Research and Evaluation, *Review of Research in Education*, Vol. 8, pp. 158-233.

BURSTEIN, L., OAKES, J. and GUITON, G. (1992), ''Educational Indicators'', *Encyclopedia of Educational Research*, Sixth Edition, Vol. 2, pp. 409-418.

CAMPBELL, D.T. (1969), ''Reforms as Experiments'', *American Psychologist*, Vol. 24, pp. 409-429.

CAPRA, F. (1982), *The Turning Point: Science, Society, and the Rising Culture*, Simon and Schuster, New York.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS (1989), *State Education Indicators*, CCSSO, Washington, D.C.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS (1990), *State Education Indicators*, CCSSO, Washington, D.C.

CREMIN, L.A. (1988), *American Education: The Metropolitan Experience 1876-1980*, Harper and Row, New York.

CRONBACH, L.J. (1982), "Prudent Aspirations for Social Inquiry", Chapter 5, in Kuskal, W.H. (ed.), *The Social Sciences: Their Nature and Uses*, The University of Chicago Press, Chicago, IL.

DE NEUFVILLE, J.I. (1975), *Social Indicators and Public Policy: Interactive Processes of Design and Application*, Elsevier Scientific Publishing, Amsterdam.

GARDNER, H. (1983), *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books, New York.

GREEN, T. (1985), "The Formation of Conscience in an Age of Technology," *American Journal of Education*, Vol. 94, pp. 1-32.

GUTHRIE, J. (1987), *Indicators of Reading Achievement*, The Rand Corporation, Center for Policy Research in Education, Santa Monica, CA.

HANUSHEK, E.A. (1981), "Throwing Money at Schools", *Journal of Policy Analysis and Management*, Vol. 1 (1), pp. 19-41.

HANUSHEK, E.A. (1986), "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature*,Vol. 24 (3), pp. 1141-1177.

KAAGEN, S.S. and COLEY, R.J. (1989), *State Education Indicators: Measured Strides, Missing Steps*, Educational Testing Service, Washington, D.C.

LINDBLOM, C.E. and COHEN, D.K. (1979), *Usable Knowledge: Social Science and Social Problem Solving*, Yale University Press, New Haven, NJ.

LIPSKY, M. (1980), *Street-level Bureaucracy: Dilemmas of the Individual in Public Services*, Russell Sage Foundation, New York.

MacRAE, D. (1988), "Professional Knowledge for Policy Discourse: Argumentation versus Reasoned Selection of Proposals", *Knowledge in Society: The International Journal of Knowledge Transfer*, Vol. 1 (3), pp. 6-24.

MURNANE, R.J. (1987), "Improving Education Indicators and Economic Indicators: The Same Problems?", *Educational Evaluation and Policy Analysis*, Vol. 9 (2), pp. 101-116.

NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION (1983), *A Nation at Risk*, U.S. Government Printing Office, Washington, D.C.

RESNICK, L. (1987), *Education and Learning to Think*, National Academy Press, Washington, D.C.

RIECKEN, H.W. and BORUCH, R.F. (eds.) (1974), *Social Experimentation: A Method for Planning and Evaluating Social Intervention*, Academic Press, New York.

RIVLIN, A.M. (1971), *Systematic Thinking for Social Action*, The Brookings Institution, Washington, D.C.

ROBINSON, J.H. (1923), *The Humanizing of Knowledge*, George H. Doran Company, New York.

ROHLEN, T.P. (1983), *Japan's High Schools*, University of California Press, Berkeley, CA.

SAXE, L. (1986), "Policymakers' Use of Social Science Research", *Knowledge: Creation, Diffusion, Utilization*, Vol. 8 (1), pp. 59-78.

SHAVELSON, R.(1987), *Indicator Systems for Monitoring Mathematics and Science Education*, Rand Corporation, Santa Monica, CA.

SMITH, M.S. (1988), "Educational Indicators", *Phi Delta Kappan*,Vol. 69 (7), pp. 487-491.

SUTHERLAND, J.W. (1973), *A General Systems Philosophy for the Social and Behavioral Sciences*, George Braziller, Inc., New York.

WEISS, C.H. (ed.) (1977), *Using Social Research in Public Policy-making*, Heath, Lexington, MA.

WEISS, J.A. (1989), "Theories of Control in Organizations: Lessons for Schools", paper prepared for the Invitational Conference on "Choice and Control in American Education" at the Robert M. LaFollette Institute of Public Affairs at the University of Wisconsin-Madison, 17-19 May.

Chapter 3

Knowledge, Ignorance and Epistemic Utility: Issues in the Construction of Indicator Systems

by

Carla Fasano
University of Wollongong, New South Wales, Australia

"Criteria for evaluating organisational effectiveness cannot be produced by some objective, apolitical process. They are always normative and often controversial, and they are as varied as the theoretical models used to describe organisations and the constituencies that have some interest in their functioning." (Scott, 1987, p. 337)

> " I understood that, when he didn't have a solution, William used to think up a number of them, very different from each other. I was perplexed.
> – But then, I ventured timidly, you are still far from a solution...
> – On the contrary, said William, I am very close, but I don't know to which one.
> – So, you don't have just one solution.
> – If I did, I would be teaching theology in Paris.
> – Do they always have the truth in Paris?
> – Not at all, said William, but they always stand staunchly by their errors.
> – And you, I said cheekily, don't you ever make mistakes?
> – Of course I do, he replied, but instead of coming up with only one kind of error, I manage to conjure up a number of them, so I am not enslaved by any." (Eco, 1980, pp. 308-309)

Education indicators in the research literature

The tables of contents and subject indexes of well-known works on the history of education do not mention education indicators, efficiency and effectiveness, or evaluation. Instead, authors such as Archer (1984), Bowen (1972-81), Mialaret and Vial (1981), Manacorda (1983), and Ravitch (1983) use different blends of pedagogy, philosophy, psychology, sociology and political science to explain which population groups have been provided over time with which types of educational provisions, and why. Despite occasional explicit statements, no particular prominence is given to the actual performance of education in relation to the stated objectives, and even less to the issue of costs. This information can be obtained only indirectly from accounts of education reform, changes in school participation and school programmes, and the like. In current indicators parlance (Scheerens, 1990*a*), these studies all tell more about inputs into education and its context than about its outputs or outcomes.

Education indicators as a contemporary phenomenon

Although in some countries education indicators do have a history as part of social indicator systems (Caplan and Barton, 1978; Scheerens, 1990*b*, p. 62), the emergence of education indicators as an autonomous and major element of educational decision-making, practice and research in most industrialised countries dates back less than a decade.[1]

In times of economic stringency, when education costs increase beyond limits affordable to the public purse and the demand for public accountability is at its greatest, indicator systems have been heralded as a means of both controlling the former and meeting the latter. However, due to the relatively recent emergence of this tool for governance and public relations in education, its effectiveness cannot yet be assessed. Until formal evaluations of education indicator systems are carried out, it may be possible to assess at least the soundness of the premises on which indicator systems are being constructed. This chapter proposes one such analysis.

The recent emergence of education indicators can be analysed from a number of different viewpoints which help to understand their current features and proposed use. And, in a field where continuity of observation and measurement – the creation of a longitudinal baseline of indicators – is paramount to their efficient use, a historical analysis can be particularly useful.

It might appear strange to focus on history for an area that has so recently emerged. Yet, assessing the soundness of the premises of indicator systems requires analysis of the past, as current configurations have not emerged from a void. Unravelling the ''story-line'' of indicators may help explain these configurations. An historical perspective may also provide arguments for evaluating the soundness of the premises of current indicator systems and for sketching a perspective for the future.

Defining the focus on information and knowledge

The lack of records of education indicators might be attributed to at least three factors: the unwillingness of authorities to produce indicators; their reluctance, having produced them, to make them available outside a restricted circle; and, less politically, the absence of appropriate paradigms.[2]

Insight into the first two factors has been provided by analyses of the political and social factors that have influenced the emergence and evolution of educational evaluation in its various forms, and by studies, carried out by both government and research agencies, on the evolution of policy and planning modes in education and on the production and diffusion of relevant structural and evaluative data (Walberg and Haertel, 1990; Nisbet *et al.*, 1985).

This chapter focuses on the third factor, *i.e.* on the information dimension of indicators, following the definition adopted by the CERI project on International Education Indicators.[3] Indicators are intended to provide information to assist with monitoring, decision-making, troubleshooting, prediction and action in the education system; the origin, nature and utility of information upon which education indicators are constructed, particularly in relation to knowledge, will be examined. Additional paradigms developed in education or adopted from other discipline areas will be studied, as will the possible evolution of such information in the near future and issues of marginal utility associated with the resulting knowledge gains.

Indicators and paradigm development in education

The concept of indicators as information appears applicable to the fields of educational evaluation, educational planning, and educational administration, all of which appear equally likely to provide an appropriate paradigmatic baseline for indicator systems. To identify the origin, nature and utility of the information on which education indicators are based, one might begin by analysing paradigm development in these three fields, followed by a study of how such paradigms are reflected in the construction of current indicators.

An early undifferentiated discipline area

At the outset, it must be emphasized that none of the three fields mentioned above existed as a recognisable study area endowed with its own body of theory, methodology and applications until well into the twentieth century. The early history of paradigm development in these areas must be traced through the development of paradigms in education and pedagogy. Until the nineteenth century, the operating reasoning mode, where it existed, had its basis in practice rather than in theory. The cornerstone of both the training and practice of educators and education administrators has long been, at best, "instrumental" rather than "cognitive" reasoning.[4]

In most industrialised countries, emphasis on applications and practice in education has been correlated with the separation of research institutions from training bodies for educators and educational administrators. The absence of educational topics from most research programmes maintained and reinforced this separation until the second half of the nineteenth century. At that time, the introduction of pedagogy into academia (Balduzzi, 1986; Husén, 1984; Bates, 1980) initiated a process which in time moved the balance between instrumental and cognitive reasoning in education away from the former and towards the latter.

Pedagogy was not admitted into academia, however, on grounds that it would allow knowledge to develop through "instrumental reasoning" from within education by using a recognised methodology. Instead, the new academic area of education was created under the administrative and conceptual tutelage of already established disciplines. The marks of this beginning are still visible in current indicator systems and, as argued below, they continue to affect the soundness of the premises of indicator systems and their ability to fulfil their function.

The early paradigm definition in education varied, depending on existing cultural differences. Thus, philosophy had a profound impact on the development of pedagogy in countries such as Germany and Italy (Husén, 1984; Balduzzi, 1986; Granese, 1986). In Sweden and the United States, early chairs in education were occupied by psychologists, particularly by experimental psychologists (Husén; Bates, 1980). In the United Kingdom and France, education came relatively late to the universities, but then also developed under the influence of psychology (De Landsheere, 1981), which by then permeated most educational research in industrialised countries.

Early educational research was quite narrowly focused on the classroom, independent from its bureaucracy, the community, and society at large. What was taught and, equally important, how it was taught, were the subjects of considerable research. Under the increasing influence of behaviourism, attention was also paid to what was actually being learnt. The almost exclusive focus on teacher and pupil, and on the process and content of teaching and learning, was the distinctive mark of "cognitive reasoning" in early education and remained the predominant perspective for several decades. It still exerts an influence on education indicators.

From the late 1800s to the late 1940s: early paradigm development in education evaluation and administration

Although forms of education evaluation were practised as far back as the second millennium BC by the Chinese (Worthen, 1990), formal and systematic evaluation studies were not carried out until the late 1800s and early 1900s. These early evaluations conducted in the United States focused on schools, teacher efficiency, and learning gains. Between 1900 and 1930 indicators were developed for students' learning performance on the basis of testing instruments – standardized and norm-referenced tests – developed within psychology paradigms; also included were a grab-bag of indicators inspired by Taylor and Fayol's theories of scientific management (Scott, 1987) which included expenditure, drop-out rates, and promotion rates. These early attempts to construct indicators and ensure their acceptance within the wider community were often plagued by accusations of "muckraking" and "propaganda". In the United States, university institutes were formed during the 1920s and 1930s; they specialised in field studies and conducted better quality surveys for local districts (Madaus *et al.*, 1983); no nationwide evaluations of education were carried out.

In the early 1930s, when social indicators were first developed on a national basis in the United States (Caplan and Barton, 1978), education evaluation came of age. The *Eight Year Study* (1932-40), funded by the Carnegie Corporation, introduced a new and broader view of education evaluation. At its core was Ralph Tyler's paradigm, in which educational evaluation was conceptualised as a comparison between intended and actual outcomes of school programmes (Madaus *et al.*, 1983). Indicators were constructed accordingly.

The knowledge base of education administration has seen a similar evolution. While the history of schools and other educational institutions, and their financing, were among the initial approaches, scientific management soon became more prevalent in the education administration curriculum (Bates, 1980).

Throughout these early developments, where paradigms from psychology, behaviourism and scientific management predominated, evaluation and administration of education were confined to the administration and evaluation of schools – their principals, teachers and students. This circumscribed focus of research continued to be predominant over subsequent decades, although its paradigmatic baseline changed substantially.

From the late 1950s to the late 1980s: contemporary paradigm development in educational evaluation, administration and planning

The late 1950s and 1960s brought about many changes in education. The predominance of psychology and philosophy in education theory and of scientific management in educational administration was challenged. Economic growth and expanding enrolments in education created a new environment for policies and programmes. As new population groups entered education, and as the industrialised world developed new national priorities, such as technological advancement and social justice, a need arose for alternative analysis and evaluation. Social, political and economic concerns came to share the stage with pedagogical issues. The school and classroom were no longer seen in isolation from the community and society at large but as a locus where social, political and economic forces merged to affect teaching and learning. The changed perspective meant that disciplines such as educational sociology, politics of education, history, and the curriculum took on greater importance. Economics of education also took on importance, although it developed more sporadically. Even as late as the 1980s, it was still not recognised as a prominent educational discipline in many OECD countries (*e.g.* Australia).

The development of the theory of human capital in the 1950s provided education with a powerful planning instrument. Models were evolved on the basis of the input-output analysis developed by Leontieff in the 1940s, itself based on the concept of production functions. These paradigms introduced the assumption that the education system was engaged in a production function by which variously defined inputs were transformed into variously defined outputs as efficiently as possible in order to maximise the input/output ratio. This has since been the most influential approach for the construction of indicator systems.

The production function approach to education has produced three broad kinds of indicators: manpower requirements, rate of social return, and social demand (Psacharopoulos, 1987). Indicators developed within models based on long-established paradigms in economics typically include input variables related to levels of resources, students and teachers, throughput variables such as transition proportions, and output variables such as numbers of graduates. Rate-of-return models add other indicators on costs and benefits related to the system itself or its clients. These models also include contextual indicators as they relate to labour market characteristics and conditions.

To assess the soundness of indicator systems built upon these models, it is as instructive to see what information they do not provide as it is to see what information they do. The choice of indicators implicitly assumes that the characteristics of the system they emphasize are more significant than others to the efficient functioning of the education system. Among the variables excluded are the legal and administrative characteristics of education, such as its centralised and decentralised nature, and its school-leaving age benchmarks. These systems also postulate, again by default, that factors such as industrial relations and the allocation of power and responsibilities among local, state and federal governments have little bearing on how effectively and efficiently education

BEST COPY AVAILABLE

systems perform. Thus, emphasis on the production function approach tends to narrow the vision in indicator systems.

During the late 1950s and early 1960s, impetus for curriculum reform in the United States and elsewhere led to major new curriculum development programmes and, subsequently, to the need for their evaluation. The conceptual and methodological impoverishment of the field soon become apparent. Worthen (1990, p. 43) notes that, despite a promising early start, "theoretical work related to educational evaluation, per se, was almost non existent", and evaluators of education programmes were left "to glean what they could from other fields". They relied to a large extent on psychology, but also applied input-output models derived from macro-economics.

The resulting almost exclusive reliance on quantitative methods was challenged in the 1970s, when qualitative methods gradually were accepted by education evaluators. While the quantitative and qualitative schools of thought were at first opposed to each other, the late 1970s saw them join together with a movement independently started in the late 1960s to help practising programme evaluators. Together, they produced evaluation models in which the two approaches could be harmonized (Worthen, 1990).

By the end of the 1980s, the influence exerted by the enlarged paradigm base of education evaluation had produced a large array of not altogether incommensurable models. Among them, social psychology models, with a focus on evaluation of curriculum and instruction, bear the most visible mark of Taylorism (objectives, established and measured behaviourally, have high priority; student performance is also defined and measured behaviourally; "what is" is compared with "what should be").

Evaluation models were also developed to meet the specific needs of decision-makers. Prominent among them were the decision models based on the CIPP model (context, input, process and product evaluation) (Borich, 1990), where input-output economic models add elements of scientific management and social psychology approaches to organisational functioning, such as those which define organisations as natural and open systems.[5]

The paradigm expansion of the 1970s and 1980s meant also that quantitative methods became less central in education evaluation models. "Softer", qualitative methodologies were in fact at the basis of most widely used models, such as judgement-oriented approaches in which experts judged programmes using a variety of methodologies. In addition, there were the adversarial and pluralistic-intuitionist approaches in which the evaluator's view is confronted and/or negotiated with those of individuals and groups served by the programme (Worthen, 1990). The latest approaches imported from organisational studies are at the basis of other evaluation models, where decision-making approaches, stakeholder analysis, and game theory are used to advantage to construct yet other models of evaluation (Scheerens, 1990*b*). And the list goes on. An inventory of actual and potential indicators constructed on these models would not only span the whole length of the quantitative-qualitative dimension, but cut across several parent discipline areas as well. Within this context of theoretical and empirical affluence, a particular tradition – that of effective schools research – has been most influential in the development of current indicator systems. It has provided clearer understanding of the "process" and "context" boxes in the indicator framework (Scheerens, 1990*a*).

The paradigmatic grounding in economics has given strong coherence to the development of models and indicators in education planning, and the absorption of paradigms from a wide array of disciplines has provided education evaluation with its distinctively eclectic but not incoherent profile. On the other hand, the evolution of models in education administration has been far more laborious.

Despite promising beginnings, the field of education administration came to a standstill in the 1950s. It was only in the late 1950s and 1960s that, under the impact of widespread paradigm enrichment in education, education administration also diversified its approaches. Paradigm expansion and field fragmentation followed in the 1960s and 1970s; since the early 1980s, there has been a much heralded renewal, which has included investments in further theory building and attempts to increase communication and coherence among sub-fields, with implications for the construction of indicators.

The focus on scientific management produced a paradigmatic base in education administration that, over time, revealed important conceptual limits (Scott, 1987). Rationality of decisions, clarity of goals, and centrality of job analysis were all questioned as major components of organisational effectiveness. A 1957 landmark publication, *Administrative Behaviour in Education*, introduced psychology, social psychology and philosophy. Sociology, political science and economics entered education administration over the following two decades, during which time the sub-specialisation of school law also took hold (Bates, 1980). By this time, education administration also encompassed concepts and activities related to the planning and management of educational organisations. The label "education administration" has this expanded meaning in the following discussion.

The expanded imports of paradigms in education administration failed to produce agreement among specialists. The separation between different imports remained and/or intensified over the years, producing a discipline which remained highly fragmented for over two decades (Bates, 1980). The comparatively low academic status assigned to education administration is the result, according to some analysts, of a lack of consensus over theoretical issues, as well as the low level of research methodology, the practical nature of the activity, and the political nature of the field (Bates; Hoy, 1982).

The gloom began dissipating in the early 1980s:

> "Both theory and research have been recently subjected to rigorous analysis and severe criticism. Although the shortcomings are numerous and the progress limited, a dynamic tension growing in the field has the potential to revitalise the study of educational administration... The present creative ferment in theory should provide researchers...with the conceptual basis to make systematic inquiry relevant to both theory and practice." (Hoy, 1982, pp. 8-9)

Policy analysis in particular has been identified as having "the potential to link theory with practice and bridge the various specialty areas in education administration" (McCarthy, 1986, p. 10). The development of the field of education administration has meant that its potential contribution to the construction of indicator systems has failed to materialise (Bryson, 1989).

The consequence of this loss in terms of the soundness of the premises of indicator systems should not be underestimated. Indicators generated by the collective instrumental and cognitive reasoning of education administrators could be among the most revealing about organisational processes in education. In addition, they might have proved very reliable in exposing influences from external quarters. At present there are very few of them in indicator systems.

If education planning has generated a reasonable number of indicators, and education evaluation possibly too many, education administration has produced far too few. The imbalance may affect the soundness of the premises upon which these systems are constructed. A brief analysis of the paradigmatic base in some of the current and/or advocated indicator systems is informative in this regard.

Paradigms reflected in current indicator systems

As indicated above, the paradigm development in education planning and evaluation has left visible marks on indicator systems, both those proposed (Scheerens, 1990*b*) and those actually adopted in some countries.[6] In both cases, the mark is most visible in the selection of indicators included under the headings.

The "outcomes", "resources" and "context" indicator framework adopted in the United States is a hybrid that contains elements of the input-output models of education, moderated by models of education evaluation. This indicator system contains indicators selected from each of the three approaches of manpower requirements, rates of social return and social demand. Other indicators are constructed from decision models belonging to the CIPP lineage. These combine elements from economic and social psychology models of organisations and appear especially under the "context" subcategories of "student characteristics" and "learning environment", which the education evaluation tradition would perhaps be more inclined to place in a separate "process" box.

Quebec's indicator system does not use the customary paradigmatic headings at all. Instead, it groups indicators under headings such as "financial resources for elementary and secondary education", "progress through school", "evaluation of learning", "secondary school graduates" and "adult education". Within these groups, however, it is not difficult to recognise some customary indicators from the same paradigms – input indicators, process indicators and output indicators – with the familiar contributions from the manpower requirements, rates of social return and social demand models. The social psychology connotation of education evaluation models is, however, absent from this indicator system which relies heavily on quantitative information.

These considerations also apply to the indicator system being developed within the OECD/CERI project on International Education Indicators[7] as well as to other variants of indicator models, such as those analysed by Scheerens (1990*a*) and van Herpen (1992). When attention is paid to the actual indicators rather than to the categorical headings or the conceptual justification of the framework, it is not difficult to trace the individual indicators to their paradigmatic roots. Models based on economics within educational planning, as well as educational evaluation models derived from models of organisations from social psychology, dominate the indicators appearing in the different indicator systems. These indicator systems also share an almost exclusive identification of schools as the focus of the effectiveness of the education system at large. This should be considered in some detail when constructing indicators.

Is this paradigmatic baseline sufficient to judge the soundness of the premises on which indicator systems are being constructed? Do the resulting indicator systems offer a comprehensive profile of the factors that determine the effectiveness of the education system? Among all the parts of the education system, is the school alone relevant to education effectiveness? Is this hypothesis validated by appropriate theory and methodology? Or is it intended as an axiom? These and related questions form the focus of the analysis hereafter.

Maximising knowledge

The questions above are phrased in terms of theories and the causal relationships they identify. They are also phrased in terms of the reasoning – instrumental or cognitive – which may lie at the origin of the paradigms used to construct current indicators. In other words, the questions are phrased in terms of the validity of these paradigms. They are also phrased in terms of coverage. Whether or not current indicators pass the validity test, the question of whether they are all the relevant indicators for assessing the education system's effectiveness remains. Under scrutiny here are the validity of the current selection of paradigms and the exclusive focus on schools as the only element of the education system to appear in the equation for education effectiveness. In short, by addressing issues of validity, coverage and focus, the questions concern the issue of the knowledge content in current indicator systems. They highlight the need to establish criteria whereby the knowledge base of indicators could be assessed in order to maximise it.

Issues of validity

In appraising the validity of theory and methodology in their disciplines, researchers have found that the production-function model of education has produced disappointing results (Scheerens, 1990*a*; Blaug, 1987). Even when students' background characteristics (*i.e.* level of intelligence and socio-economic background) are taken into consideration, input indicators (*i.e.* teachers' salaries, experience and qualifications, teacher/pupil ratios, and per pupil expenditure) have not shown clear correlations with output indicators such as student achievements. Only one indicator, teacher experience, has shown some degree of correlation across studies (Scheerens, 1990*a*).

From these results, commentators conclude that the inputs so identified are not linked causally to the outputs, and that there is little point in expecting that an increase in an input such as expenditure, for instance, would increase education effectiveness, or that policy measures to this effect would be warranted. They are quick to dismiss the fact that the lack of a clear correlation between inputs and outputs could also be explained quite differently. Blaug, a prominent education economist, suggests (1980) that the poor correlation between input and output indicators could be a result of the poor validity of the paradigms.

In Blaug's opinion, several epistemological weaknesses plague the discipline of economics. Among the diagnosed ills are: measurement without theory; replacement of falsification with verification tests, which, as a good Popperian, Blaug deems a weaker test of the value of theory; the emergence of new approaches by radical political economists who still rely on old measurements to support new explanations; and the latter-day Austrian reaction against empirical methods. Neo-classical economists, in particular, who have had the major influence on education planning approaches, are accused of not practising what they preach in terms of methodological rigour, and producing information lacking in predictability and policy significance.

Criticism of the validity of paradigms also abounds in the field of education evaluation. The lack of an adequate knowledge base for programme evaluation ranks high among the issues identified. Others are the scarcity of empirical information about the relative efficacy of alternative evaluation models and the extent to which various data collection techniques interfere with ongoing educational processes. The lack of clarity on the utility of evaluation combines with the rarity of meta-evaluation (assessment of evaluation studies) to further weaken the foundations of education evaluation. More specific criticism of the decision-oriented evaluation models, which have strongly influenced the construction of indicators, points to their unfounded assumption concerning the existence of clearly identifiable decision-makers and their lack of awareness of the political nature of evaluation (Worthen, 1990).

Issues of coverage

It is apparent that the current lists of indicators are incomplete. For instance, the field of education administration could undoubtedly have facilitated the construction of indicators on those political factors which often make or break education reform. It could also have constructed indicators on the formal or informal support or hindrance by vested-interest groups to the efficient management of an education system or of a single school. It could have constructed indicators to identify possible differences among the determinants of effectiveness in centralised or decentralised education systems. Finally, through its sub-field of school law, it could have developed indicators on the legal framework, which either constrains or facilitates the production function of schools as well as their socialisation function.

The above list is indicative of the kind of indicators that could be constructed on the basis of available knowledge about the management and governance of the education system. Other contributions might come from other disciplines; for one example among many, education philosophy could propose indicators pertaining to legitimacy issues that shape policy formation and influence its efficient implementation (Walker, 1991).

Evidence suggests that the currently available indicators have not always gone through rigorous tests of relevance, reliability and validity. Rather, the range of paradigms represented appears to have been defined by default. The groups of professionals and researchers who, at critical times, began to focus on issues of effectiveness in education simply represented a restricted number of paradigmatic perspectives. As the traditions they began were consolidated, other paradigms, including some which are potentially highly relevant, remained peripheral and still fail to influence the construction of indicators in any substantive way.

Arguments that the paradigm base of indicator systems needs to be expanded might well be challenged. Incumbents seldom accept wholeheartedly efforts to open up territorial boundaries, unless, of course, they are taking the initiative themselves. This is often the case, and the most telling arguments in support of paradigm expansion in indicators may be coming from those education fields in which current indicator systems have been constructed.

Since the early 1980s, the fields of education planning and education evaluation have identified their limits and proposed paradigm expansion as an important remedial strategy. In his critical account of the economics of education, Blaug (1987) concludes that the new generation of economists of education is likely to improve theory and overcome the proven weaknesses of the production function approach to education by importing additional paradigms that account for socio-psychological variables in the school and the workplace and include more detailed knowledge of labour market behaviour.

In his critical analysis of school-effectiveness research, Scheerens (1990*a*) also proposed solutions involving paradigm expansion which should make it possible to produce education evaluation models directed to the support of administrative and political decision-making. Among the models proposed are evaluations focused on utilisation, evaluations based on stakeholders, evaluation according to betting models, and active "rational reconstruction" by evaluators. These models attempt to achieve more comprehensive conceptualisation of the context of evaluation practice, by taking explicit account of organisational and institutional arrangements.

The proposed paradigm expansion in education evaluation is also reflected in arguments made concerning the political nature of education evaluation. Commenting on the "image of command" reflected in certain models and on the reality of decisions taken *de facto* by many of the actors who form the "policy-shaping community", Cronbach and his associates enlarge the previous list of actors in education evaluation, which comprised programme monitors and project personnel, to include policy-makers, bureaucrats, vested interest groups and the like (Cronbach *et al.*, 1980).

The focus on schools

If and when it occurs, a paradigm expansion in indicator systems is likely to challenge a major assumption: that schools, isolated from the bureaucracy of the system, are the most relevant element for determining the effectiveness of the education system. Until very recently, this assumption had never been properly tested, either to verify or falsify it, and its examination is long overdue. The examination might reveal factors that are at least as influential as school-based ones.

The reasons for the exclusive focus on schools and the neglect of the education bureaucracy are to be found in the same historical events that led certain paradigms to predominate. Because of the paradigmatic foundations of pedagogy in philosophy and psychology, enquiry focused on the nature and operation of schools and tended to ignore the bureaucratic structures. This focus has seldom been shifted or expanded by subsequent cognitive reasoning and theory-building to include other aspects of the education system, even in countries where a centralised system suggests very strong rather than loose coupling of the schools and the educational bureaucracy.

The paradigms that underlie current indicator systems reflect the same tradition. They either give little importance to the organisational design of the education system and represent it as a black box (educational planning), or they centre on outcomes of school programmes (school evaluation). Even models of planning and evaluation that focus on student characteristics reduce the educational organisation to those locations where students are found; it is in the schools that educational provisions, decided upon elsewhere in the organisation, materialise.

History may again be of assistance in explaining why education bureaucracy has been excluded from the planning and evaluation of education. One reason can be found in the fact that the geographical source of most of the original paradigms and the bulk of education research relevant to indicator systems is the United States and the United

Kingdom, countries with strongly decentralised education systems, where the idea of the autonomous school is a cherished value.

Researchers in these countries have naturally tended to conduct their work within their domestic organisational structure and in obedience to its "school autonomy" ethos. When they have looked inside the black box of the education system, they have tended to overlook factors for school effectiveness that go against the ethos, those that lie beyond the school and its immediate community. When decision-making, administration and management not based on the school have been included, they have, at best, produced a few indicators, confined to the "context" box alongside other "community characteristics" (van Herpen, 1992; Scheerens, 1990*b*). These research traditions have been uncritically adopted elsewhere, despite structural, administrative and legal differences in education systems. Hence, the generally shared belief that education effectiveness means the effectiveness of schools and that, by default, education effectiveness is not related to the effectiveness of education bureaucracy.

To paraphrase Thomas Kuhn, this shared belief is the power behind the notion that schools are the "locus of professional commitment" of educators and education administrators, a notion held "prior to the various concepts, laws, theories and points of view that may be abstracted from it" (Kuhn, 1970, p. 11). It also authorises the assumption that "as long as a belief whose causes are undetected is not challenged by other persons, and engenders no conflict that would prompt us to wonder about it ourselves, we are apt to go on holding it without thought of evidence" (Quine and Ullian, 1978, p. 15). Alternatively, it could be argued that, given the original absence of a focus on bureaucracy, the belief has prevailed as a manifestation of the principle of sufficient reason, which holds that justification is sought for what is there and not for what is not there (Taylor, 1983), hence the constant improvement of school-focused theory of educational effectiveness and its applications and the absence of equivalent studies of the bureaucracy.

Empirical evidence that bureaucracy does count in determining the effectiveness of education and, in particular, the effectiveness of schools, has begun to emerge. Some of it, most telling, comes from decentralised education systems such as that of the United States. Kirst (1989) has documented the increasing interference of non-school-based governance structures in the management of schools. The trend, brought about and consolidated over the last 20 years by increased reliance on non-local funding of schools, has currently reached – according to Kirst – unacceptable levels. It has supported the diversification and specialisation of school functions and, in addition, has introduced a body of school specialist personnel more loyal to their distant sources of funding than to the local school authority. Increased influence from private interest groups and professional reformers also tends to weaken the policy-making capacity of superintendents and school boards. Non-local influence on schools' curriculum, performance, and behavioural standards has been monitored during the past years, and the findings indicate that it may have negative effects on school effectiveness (Kirst).

Chubb and Moe (1990) have produced quantitative estimates on the negative effects of non-school influences on school performance.[8] In his introduction to the study, B. MacLaury, President of the Brookings Institution, aptly summarised its main finding: "Government has not solved the education problem because government is the problem" (p. ix). Briefly, Chubb and Moe define bureaucracy as the superintendents and the central office administration and show that their influence over the hiring and firing of teachers, the curriculum, instruction and discipline does indeed affect the efficient organisation of schools. Influence on hiring and firing of teachers is found to be the strongest factor, followed by that on curriculum and instruction. The interference of teachers' unions in hiring and firing also correlates strongly with school effectiveness. Student body characteristics were found to be only the second most important factor.

Over a four-year high school career, identical students attending effective and ineffective schools would differ by more than a full year in achievement gains. The influence that the bureaucracy exerts over school organisation is sufficiently strong that it alone is capable of producing most of this achievement difference (Chubb and Moe, 1990, p. 165).

These results may seem radically new when compared with the traditional studies of school effectiveness which have focused on elements internal to the schools or, at best, its most immediate community. They correspond, however, to what organisation and policy studies anticipate for the public sector and can offer ways to expand indicator systems.

The field of policy studies would focus the construction of indicators of education effectiveness on paradigms related to policy implementation, which is what schools are mostly involved with. Research has consistently shown that particularly influential factors for success in this area include those related to the policy's information base: validity of the identification of the problem addressed by the policy and of the assumptions/theories of causes and effects underpinning the proposed solution, as well as the understanding of the policy objectives by all concerned. Other factors relate to the policy's resource base and to its adequacy in terms of financial, human and time resources and their combination. The communication base of policy is also important to successful implementation and includes efficiency of communication channels and numbers of critical decision-making points in the process of implementation. When education is considered as a public sector organisation like any other, this includes both its bureaucracy and its delivery points, the schools.

Successful policy implementation is critically dependent upon the nature and configuration of its power base, *i.e.* on the importance of (political, pedagogical) stakes, on consensus on the policy's objectives, and on control over circumstances affecting implementation, as well as the capacity to demand and obtain compliance. The rate of return to the implementers and their clients, the professional competence of implementers and the degree of co-ordination between the policy and other relevant policies are also important (Hogwood and Gunn, 1984). Few of these factors are reflected in current indicator systems. Most are clearly related to the issues raised by the educational research reported above.

Knowledge from the field of organisational studies would orient the expansion of indicator systems in the same direction. Seen from this vantage point, the education system can be characterised as a combination of the types of organisation identified by Scott (1987) as rational, natural and open systems (see note 5). Seen as a rational system, it is an organisation characterised by high goal specificity and highly formalised social structures, with clear lines of responsibility and accountability, whose officials constitute its bureaucracy and work either directly or indirectly to achieve its goals (Scott). Seen as a natural system, it has collectivities that engage in informally structured activities to secure common ends. Considered as an open system, education is a coalition of shifting interest groups that develop temporary common goals by negotiation under strong influence from the environment.

In addition to the "rational" indicators which currently dominate indicator systems, especially in the "input" and "output" evaluation boxes, "natural" and "open" indicators could be constructed (Scheerens, 1990*b*). There are a small number of "natural" indicators in some indicator systems, mostly in the "process" box (leadership, school climate, expectations, motivation); others could focus on the complexity of the system's goals, and in particular, the match or mismatch between stated and real goals pursued by all participants, in schools as well as in the bureaucracy. "Open" indicators, now absent from most indicator systems, would focus on the throughput of resources fbetween the education system (schools and bureaucracy) and its environment, and on the actual flow of information, resources, materials and the like within the system, a flow which may or may not match that of the organisation's formal chart. They would also focus on patterns of allocation of power and accountability throughout the system and its environment.

Research from an organisational perspective could add further details to indicator systems. There are few examples of indicators on the sharing of management responsibility and control between managers and professionals at different levels of seniority (*e.g.* teachers and less senior bureaucrats), sharing that would help ensure effectiveness in a professionalised education system (Benveniste, 1987). Indicators on the identification of the organisation's mission, vision, internal strength and weaknesses, external threats and opportunities, would be central to strategic planning in the education system (Bryson, 1989).

These approaches, among others, could be analysed with a view to identifying education indicators that would allow better informing, monitoring, troubleshooting and predicting in the system. Most of these indicators would have to be included in the "input" and "process" boxes, particularly those dealing with the functioning of the bureaucracy, internally as well as at the interface with the schools.

It is quite clear that these additional indicators would improve the knowledge base of indicator systems. It is also beyond doubt that, as the paradigms currently represented improve their internal knowledge base, new indicators would replace current ones. It may therefore be safely assumed that if maximisation of the knowledge base of indicator systems is pursued with determination and consistency, the production of indicators could well become an endless process in which transformations would alternate with additions. This would, of course, be very costly. Even if the epistemological and political obstacles were removed, the economic constraints would remain. Can countries afford to produce a potentially endless list of indicators? What benefits are to be obtained in return? Are indicator systems to become a bottomless pit of public resources? Can the length of the list be kept compatible with what the public purse can afford? Can the political risk of operating with indicators unlikely to provide all the information needed to maximise the system's effectiveness be taken? Should governments design some alternative strategies to knowledge maximisation? The following section proposes some possible options.

Managing ignorance

The notion that information is a resource, and a costly one at that, has been discussed since the 1960s by economists, who have applied to it the concept of utility (Machlup, 1980), and by philosophers, who have developed the concept of epistemic utility, where knowledge rather than information is at stake (Hempel, 1965; Mattessich, 1978). If knowledge is a resource, then the law of diminishing returns may be expected to apply; marginal increases will yield improved benefits, but only to a point: beyond a certain level, the rates of return will begin to decline.

Given the organisational and political stakes involved in constructing indicator systems, the importance and urgency of developing definitions and estimations of their epistemic utility are obvious. In particular, in view of the costs involved, it is important to estimate when and how, if governments continue to fund the collection of better and longer lists of indicators, the law of diminishing returns would apply.

The fundamental dilemma in the construction of education indicators is that complete and documented knowledge of the education system of the kind needed to construct scientifically valid indicator systems is virtually unattainable. This dilemma has been presented in this chapter from the point of view of the availability and application of theory. In addition, political realities, time constraints, speed of change, and costs also rank high among the factors hindering the maximisation of paradigm-based knowledge.

For all these reasons, indicator systems will have to incorporate at all times a certain amount of non-knowledge. The risks run by education decision-makers when they use indicator systems, especially the risk of taking them as an optimum system, should not be underestimated (see Frackman, 1987). A main danger is that they will reorient education towards the organisational models implicit in current indicator systems, primarily the rational model of scientific management. Organisational science has long seen this model as leading to ineffective management of organisations.

Specialised and highly trained developers and users of indicator systems are aware of this problem as well as others that limit the validity of indicator systems. These experts can and do avoid these limitations by taking them into account in their deliberations and recommendations. But what kind of analytical capacity can be expected from the average user of indicators in central or regional bureaucracies and local school settings, from journalists, or from a parent?

These decision-makers and users reach conclusions, form opinions and act on the basis of information and data for which knowledge is often indistinguishable from "non-knowledge". Knowledge-oriented strategies to prevent, or at least limit, misjudgement, might include specialist training and community education campaigns designed to confirm the central value to education that the construction of its indicator system represents.

Recent trends in theory-building question the classical assumption that "knowledge is to be sought and ignorance shunned" and propose a more positive approach to the latter (Smithson, 1989). This approach might be used as a guide to constructing indicators that take account of who needs access to which indicators and for what purpose, what information and knowledge can and should be overlooked in constructing indicator systems, how much users really need to know, and how much non-knowledge and of what kind decision-makers and users of indicators can really afford.

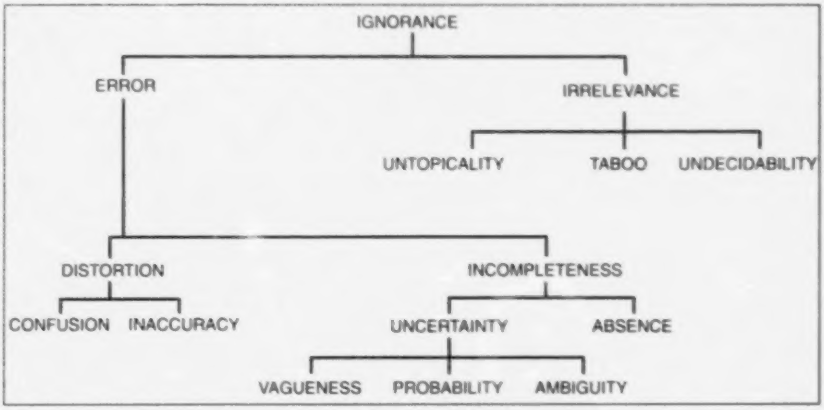
The first step towards answering these questions is to clarify the meaning of "non-knowledge". From an economist's point of view, Machlup (1980) defines "non-knowledge" as "negative knowledge" and includes: knowledge that is unwanted, not comprehended or only partially comprehended; knowledge that is restricted or forbidden, disproved or suspended; or knowledge that is losing relevance or is simply confusing.

From an epistemologist's standpoint, Smithson (1989) subsumes "non-knowledge" under the comprehensive heading of "ignorance" and proposes a more comprehensive taxonomy. If "A is ignorant from B's viewpoint if A fails to agree with or show awareness of ideas which B defines as actually or potentially valid" is accepted as a broad definition of ignorance, then it is useful to recognise that ignorance can mean different kinds of "non-knowledge" calling for different management strategies. According to the proposed taxonomy of ignorance, it is appropriate to distinguish between irrelevance, ignoring or overlooking, deliberately or not; error, connoting distorted or incomplete knowledge; untopicality, referring to the intuitions people have and negotiate with others; undecidability, relating to matters people are unable to designate as true or false; taboo, alluding to socially enforced irrelevance; distortion, referring to distortion in degree (inaccuracy) or in kind (confusion); and incompleteness, either in kind (absence) or in degree (uncertainty). Examples of applications of the "taxonomy of ignorance" to the construction of indicator systems are given in Figure 3.1.

The first application relates to the kind and degree of "ignorance" of potential indicators. Depending on the type of ignorance, indicators might be included or excluded from indicator sets on the basis of specific criteria. Candidates for exclusion would be indicators marked by "irrelevance" and its sub-cases of "untopicality", "taboo" and "undecidability". Candidates for inclusion are those marked by "error", provided the error is not of the "confusion" and "inaccuracy" type. Among the "incomplete", and thus acceptable, indicators are those defined within a validated theoretical framework and/or experience, possibly as an application of known cause/effect relationships. These indicators would either be marked by a degree of "uncertainty" or, if associated with "absence", be the object of further theory-building or experimentation. "Uncertain" indicators would be processed differently, depending on their relation to "vague", "probable" or "ambiguous" information, but would nonetheless be used.

"Uncertainty" has already been demonstrated to be beneficial in some policy processes. In his analysis of policy processes, Schulman (1988) claims that when a combination of vague core concepts and formalised long deductive chains are encoun-

Figure 3.1 **Taxonomy of ignorance**



Source: Smithson (1989), *Ignorance and Uncertainty*, New York: Springer-Verlag, p. 9.

tered in implementing policy, then these can be a source of considerable policy instability. In such cases, indicators that are themselves "ambiguous" or "undecidable" could be used to advantage, since "both analytical ambiguity and inconsistency could be a potential buffer state which protects against conditions under which analysis could destabilize policy" (Schulman, 1988, p. 288).

Central to the construction of indicator systems is their relative nature. Indicators are supposed to be used in comparative frameworks to provide information as to how well the education system is performing. Given the fact that there are as many definitions of "good performance" as there are education systems and stakeholders, the above criteria for inclusion or exclusion of indicators should be applied for each legitimate definition of "good performance" of the education system.

Just as central to the construction of indicator systems as various kinds of "non-knowledge" is the identification of their users. Countries' education systems vary considerably with respect to centralisation and school autonomy, and it would be difficult to develop a system of indicators that would be equally useful to all. Centralised systems require that both macro and micro information be made available to their "central" decision-makers for transformation into instructions for educational establishments. Decentralised systems instead have a system of decision-making that is biased towards micro information, and aggregated data may not be very useful to school districts or individual schools.

Indicators included as relevant at the level of a nation or state may become irrelevant at the smaller scale, and *vice versa.* Distortion could affect aggregated indicators used locally or centrally more easily than disaggregated ones produced and used locally and checked against direct experience of the local reality. Incomplete disaggregated indicators can also be completed more easily at the local level than aggregated ones. In general, it might be claimed that the margins of "acceptable ignorance" that can be afforded at the aggregate level for large-scale decisions are narrower than those for the small-scale and more disaggregated decision-making, where corrections can be tailored to suit. As a recent study contends, "knowledge should be sought only if it is needed; it should therefore be linked to functions to be performed" (Kogan, 1990, p. 237), provided, however, that the functions considered include informal as well as formal ones and take into account the changes brought to the knowledge needs of less senior personnel and less central administrators by the altering balance of power and accountability in education and the modified environment of its stakeholders (Fasano, 1991).

The preceding discussion should be taken as indicative of the kind of contribution a focus on ignorance could provide towards the construction of indicator systems and the estimation of their epistemic utility. No detailed formulae can be provided on how best to identify and manage ignorance in indicators. The task of drawing general and comprehensive conclusions about the problem must be postponed until more valid information is available. The sharing of ignorance is today's reality. The sharing of less ignorance could be tomorrow's.

Conclusion

The question posed at the outset of this chapter, "Are the premises of current indicator systems sound?", can now be tentatively answered.

The analysis of the evolution of current indicator systems from an historical and epistemological perspectives has identified three major issues. The first is that of the "paradigm expansion". It has been shown that paradigms which have proved highly beneficial to other fields of education have had a different impact on the development of indicator systems, as paradigms adopted for indicator systems have generated a narrow-minded view of how to conceptualise the determinants of education effectiveness. Education planning and evaluation approaches appear to address only part of the education effectiveness equation, to the exclusion of other, potentially relevant factors. The historical analysis has, moreover, highlighted the fact that current paradigms in indicator systems depend more on historical accident than on valid screening of competing approaches. The analysis suggests that current indicator systems should be promoted with caution as sources of sound and comprehensive information on the "health" of education systems. Considerations as to the validity and coverage of paradigms currently represented in indicators are at the origin of this conclusion.

Closely linked to the "paradigm expansion" question is the issue of "school focus". Current indicator systems, it has been argued, identify schools as the only part of the organisation of education that is relevant to determining education effectiveness. In doing so they ignore the possibly sizeable contribution, positive and negative, of education bureaucracy, not only to the effectiveness of the education system as a whole but to the effectiveness of the schools themselves. Neglect of these issues does not help the cause of education effectiveness but only perpetuates the current situation of "lit rooms and darkened corridors" in the house of education, where much is known about the functioning of schools and comparatively little about how the rest of the organisation functions.

The third main issue is the "knowledge dilemma". Complete and rigorous knowledge of the education system, of the kind that would allow the construction of effective indicator systems, cannot be attained for epistemological as well as economic reasons. Means therefore need to be found to manage the imperfect knowledge affecting current indicator systems and perhaps to develop management strategies of "non-knowledge". Such strategies are likely to be more effective, and might be less costly, than those aimed at maximising knowledge.

Together, these issues suggest that ensuring effective education and meeting legitimate demands for accountability from within that environment with the help of indicator systems is a task as complex as most experts recognise, and possibly even more so. Incremental development rather than bold changes may be called for. This would allow concomitant evaluation studies to provide better feedback to designers. A joint international effort to construct indicators effectively and efficiently, with a view to maximising their benefits and minimising their cost would clearly be a useful endeavour. Moreover, the clearer and better documented the objectives, performance and outcomes of education, the more accurate will be the targeting of further improvement, internally as well as externally, to this important field.

Notes

1. The 1983 report by the United States National Commission on Excellence in Education, *A Nation at Risk*, is commonly identified as the landmark publication at the origin of the current concern for education quality, which has in turn encouraged the development of indicator systems.

2. "Information" and "knowledge" are defined in this chapter in relation to the concept of paradigm, as understood by Kuhn (1970, pp. viii and 10), *i.e.* the "universally recognised scientific achievements... [that] provide model problems and solutions to a community of practitioners". Knowledge is restricted to categorical data and propositional statements generated within paradigms. Information is defined more broadly as those data and statements which might or might not be generated within paradigms.

3. Education indicators are defined by Bottani and Walberg (1992, pp. 14-15) as:
   - Information that describes the system's performance in achieving desired educational conditions or results. This information helps to describe the system's current functioning and effectiveness.
   - Information about features of the system known to be linked with desired conditions and outcomes. This information can help policy-makers, educators, and the public predict future performance.
   - Information that describes enduring features of the system. This information helps policy-makers and educators better understand how the system works and to assess the implication of changes over time.
   - Information that is relevant to education policy. This information provides insight into current or potential problems that are of particular concern to policy makers or that are susceptible to change through action.

4. "Instrumental reasoning" is based on the assumption that truth (true knowledge) is what is seen to achieve given objectives more often than any other available alternative. "Cognitive reasoning" assumes that truth is supported by evidence certified as valid by rigorous but arbitrary criteria such as "scientific method" (Mattessich, 1978, pp. 13-15). The categorical data and propositional statements generated by the two modes of reasoning overlap, but do not necessarily coincide with, those defining information and knowledge respectively.

5. As organisational studies evolved over time, the original rational model – central to the scientific management approach and based on a definition of organisations as "collectivities orientated to the pursuit of relatively specific goals and exhibiting relatively highly formalised social structures" – was replaced by emerging new models which have now been consolidated in recognisable schools. These include models of organisations as natural systems, *i.e.* as "collectivities whose participants share a common interest in the survival of the system and who engage in collective activities, informally structured to secure this end". Paradigm imports from social psychology have been pivotal to the development of these models. Since the 1960s, models of organisations as open systems have also emerged. They see organisations as "coalitions of shifting interest groups that develop goals by negotiation; the structure of the coalition, its activities, and its outcomes are strongly influenced by environmental factors". Paradigms

from system theories and information science, as well as social psychology, have been instrumental in developing these models. Contingency models of organisations, widely adopted in education evaluation and in research conducted by effective schools, belong to this category (Scott, 1987, pp. 20-23).

6. The indicator systems already in use considered in this section are those contained in publications of the Direction générale de la recherche et du développement in Quebec (1989) and of the National Center for Education Statistics (1989).
7. Information on indicator systems from the INES project have been listed in the project newsletter, *INES-NEWS*, which has been published since November 1988.
8. The empirical research of Chubb and Moe provides confirmation of the present author's analysis of the role of government structures in determining school effectiveness, in light of the historical and theoretical considerations reported here. The convergence of the two analyses provides further support to the claim that the role of educational bureaucracy structures in the effectiveness of schools should be the object of systematic research.

Bibliography

ARCHER, M.S. (1984), *Social Origins of Educational Systems*, Sage, London.

BALDUZZI, G. (1986), "La cattedra di pedagogia nell'Università di Bologna dal 1869 al 1985", Scuola e città, Vol. 7, pp. 288-294.

BATES, R.J. (1980), "Educational Administration, the Sociology of Science, and the Management of Knowledge", *Educational Administration Quarterly*, Vol. 16 (2), pp. 1-20.

BENVENISTE, G. (1987), *Professionalizing the Organization*, Jossey-Bass, San Francisco, CA.

BLAUG, M. (1980), *The Methodology of Economics*, Cambridge University Press, Cambridge.

BLAUG, M. (1987), *The Economics of Education and the Education of an Economist*, New York University Press, New York.

BORICH, G.D. (1990), "Decision-oriented Evaluation", in Walberg, H.J. and Haertel, G.D. (eds.), *The International Encyclopedia of Educational Evaluation*, Pergamon Press, Oxford, pp. 31-35.

BOTTANI, N. and WALBERG, H.J. (1992), "What are International Education Indicators For?", in *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, pp. 7-12.

BOWEN, J. (1972-81), *A History of Western Education*, Vols. 1-3, Methuen, London.

BRYSON, J.M. (1989), *Strategic Planning for Public and Nonprofit Organizations*, Jossey-Bass, San Francisco, CA.

CAPLAN, N. and BARTON, E. (1978), "The Potential of Social Indicators: Minimum Conditions for Impact at the National Level as Suggested by a Study of the Use of Social Indicators", *Social Indicators Research*, Vol. 5, pp. 427-456.

CHUBB, J.E. and MOE, T.M. (1990), *Politics, Markets and America's Schools*, The Brookings Institution, Washington, D.C.

CRONBACH, L., AMBRON, S., DORNBUSCH, S., HESS, R., HORNIK, R., PHILLIPS, D., WALKER, D. and WEINER, S. (1980), *Towards Reform of Programme Evaluation: Aims, Methods and Institutional Arrangements*, Jossey-Bass, San Francisco, CA.

DE LANDSHEERE, G. (1981), "La pédagogie expérimentale", in Mialaret, G. and Vial, J., *Histoire mondiale de l'éducation*, Presses universitaires de France, Paris, pp. 485-500.

DIRECTION GÉNÉRALE DE LA RECHERCHE ET DU DÉVELOPPEMENT (1989), *Indicateurs de l'enseignement aux niveaux primaire et secondaire*, ministère de l'Education, Montreal, Quebec.

ECO, U. (1980), *Il nome della rosa*, Bompiani, Milan.

FASANO, C. (1991), "Policy Education and Policy Research", in Fasano, C. and Winder, B. (eds.), *Education Policy in Australia: Practices and Perceptions*, Wollongong University Press, Wollongong, New South Wales.

FRACKMAN, E. (1987), "Lessons to be Learnt from a Decade of Discussions on Performance Indicators", *International Journal of Institutional Management in Higher Education*, Vol. 11 (2), pp. 149-162.

GRANESE, A. (1986), "Che cos'è la pedagogia? Un dibattito tra studiosi italiani", *Scuola e città*, Vol. 7, pp. 273-85.

HEMPEL, C.G. (1965), *Aspects of Scientific Exploration and Other Essays in the Philosophy of Science*, Collier and Macmillan, London.

HOGWOOD, B.W. and GUNN, L.A. (1984), *Policy Analysis for the Real World*, Oxford University Press, Oxford.

HOY, W.K. (1982), "Recent Developments in Theory and Research in Educational Administration", *Education Administration Quarterly*, Vol. 18 (3), pp. 1-11.

HUSÉN, T. (1984), "Research and Policy Making in Education: An International Perspective", *Educational Researcher*, Vol. 13 (2), pp. 5-11.

KIRST, M. (1989), "Who Should Control the Schools? Reassessing Current Policies", in Sergiovanni, T.J. and Moore, J.H. (eds.), *Schooling for Tomorrow: Directing Reform to Issues That Count*, Allyn and Bacon, Boston, pp. 62-68.

KOGAN, M. (1990), "Fitting Evaluation with the Governance of Education", in Granheim, M., Kogan, M. and Lundgren, U. (eds.), *Evaluation as Policy Making*, Norwegian University Press, Oslo, pp. 234-246.

KUHN, T.S. (1970), *The Structure of Scientific Revolutions* (2nd edition), The University of Chicago Press, Chicago.

MACHLUP, F. (1980), "Knowledge: Its Creation, Distribution and Economic Significance", in Machlup, F. (ed.), *Knowledge and Knowledge Production*, Vol. 1, Princeton University Press, Princeton, NJ.

MADAUS, G.F., STUFFLEBEAM, S. and SCRIVEN, M.S. (1983), "Program Evaluation: An Historical Overview", in Madaus, G.F., Scriven, M.S. and Stufflebeam, S. (eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, Kluwer-Nijhoff Publishing, Boston, pp. 3-22.

MANACORDA, M. (1983), *Storia dell'educazione*, Edizioni RAI, Turin.

MATTESSICH, R. (1978), *Instrumental Reasoning and Systems Methodology*, Reidel, Dordrecht.

McCARTHY, M.M. (1986), "Research in Educational Administration", *Education Administration Quarterly*, Vol. 22, (3), pp. 3-20.

MIALARET, G. and VIAL, J. (eds.) (1981), *Histoire mondiale de l'éducation*, Vols. 1-4, Presses universitaires de France, Paris.

NATIONAL CENTER FOR EDUCATION STATISTICS (1989), *Education Indicators*, Office of Educational Research and Improvement, U.S. Department of Education, Washington, D.C.

NISBET, J., MEGARRY, J. and NISBET, A. (eds.) (1985), *Research, Policy and Practice*, Kogan Page, London.

PSACHAROPOULOS, G. (1987), "Economic Aspects of Educational Planning", in Psacharopoulos, G. (ed.), *Economics of Education: Research and Studies*,Pergamon Press, Oxford, pp. 311-315.

QUINE, W.V. and ULLIAN, J.S. (1978), *The Web of Belief* (2nd edition), Random House, New York.

RAVITCH, D. (1983), *The Troubled Crusade: American Education: 1945-1980*, Basic Books, New York.

SCHEERENS, J. (1990*a*), "School Effectiveness Research and the Development of Process Indicators of School Functioning", *School Effectiveness and School Improvement*, Vol. 1 (1), pp. 61-68.

SCHEERENS, J. (1990*b*), "Beyond Decision-oriented Evaluation", in Walberg, H.J. and Haertel, G.D. (eds.), *The International Encyclopedia of Educational Evaluation*, Pergamon Press, Oxford, pp. 35-40.

SCHULMAN, P.R. (1988), "The Politics of Ideational Policy", *Journal of Politics*, Vol. 50 (2), pp. 263-291.

SCOTT, W. (1987), *Organizations: Rational, Natural and Open Systems* (2nd edition), Prentice Hall, Englewood Cliffs, NJ.

SMITHSON, M. (1989), *Ignorance and Uncertainty: Emerging Paradigms*, Springer-Verlag, New York.

TAYLOR, R. (1983), *Metaphysics* (3rd edition), Prentice-Hall, Englewood Cliffs, NJ.

van HERPEN, M. (1992), "Conceptual Models in Use for Educational Indicators", in *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, pp. 25-51.

WALBERG, H.J. and HAERTEL, G.D. (eds.) (1990), *The International Encyclopedia of Educational Evaluation*, Pergamon Press, Oxford.

WALKER, J. (1991), "Legitimacy in Educational Policy", in Fasano, C. and Winder, B. (eds.), *Education Policy in Australia: Practices and Perceptions*, Wollongong University Press, Wollongong, New South Wales.

WORTHEN, B.R. (1990), "Program Evaluation", in Walberg, H.J. and Haertel, G.D. (eds.), *The International Encyclopedia of Educational Evaluation*, Pergamon Press, Oxford, pp. 42-47.

BEST COPY AVAILABLE

Chapter 4

Choosing Indicators

by

Desmond Nuttall
London School of Economics, United Kingdom

The aim of this chapter is to examine the factors that influence the selection of particular indicators as components of an indicator system, and to derive a general set of principles for making the selection process more systematic. The chapter therefore starts with a clarification of the term "indicator", and then considers what may be learnt from the history of indicator systems in other fields. The chapter then looks at the major factors that govern the selection process and at how they have been embodied in lists of criteria proposed by workers in the field, before proposing such a set for use with education indicators.

What are indicators?

It is generally agreed that indicators are designed to provide information about the state of an education or social system. They act as an early warning device that something may be going wrong, much as the instruments on the dashboard of a car alert drivers to a problem or reassure them that everything is functioning smoothly. A dial pointer moving into the red zone is only a symptom of some malfunction; further investigation is needed to establish the cause. Viewed as reassuring or warning devices, indicators conform to the dictionary definition: "that which points out or directs attention to something" (the *Oxford Dictionary* definition; quoted by Johnstone, 1981, p. 2). If something is wrong, the indicators themselves do not provide the diagnosis or prescribe the remedy; they are simply suggestive of the need for action.

The consensus over the broad purpose of indicators does not extend to the precise definition of what an indicator is. Some authors reserve the definition to a narrowly quantitative one. For example, Johnstone (1981) observes:

> "A third feature of an indicator is that it is something which is quantifiable. It is not a statement describing the state of a system. Instead it must be a real number to be interpreted according to the rules governing its formation." (p. 4)

Others take a much wider view and would include descriptive or even evaluative statements within the scope of indicators (*e.g.* CIPFA, Chartered Institute of Public Finance and Accountancy, 1988). Almost always, though, even the broadest definition limits the concept to information and excludes analysis or discussion.

Those who adopt the wider view fear that limiting the concept to the quantitative will mean that indicators will be unable to portray the full richness and diversity of the educational process, and that, at worst, they will merely indicate the trivial and focus attention on the unimportant. Those who espouse the qualitative approach make a similar criticism of quantitative research in education and the social sciences. Those who propose indicators must therefore demonstrate that they are not too reductionist and will not divert attention from other important goals.

The commonest view of indicators seems to be a quantitative one. For example, in the survey carried out under the OECD Institutional Management in Higher Education programme, an indicator is defined as "a numerical value" and the OECD Education Indicators Project has tacitly taken the same view. Some authors suggest that indicators imply a comparison against a reference point (as in a time series or an average) while statistics do not; but in fact it is rare that the interpretation even of descriptive statistics dispenses with comparison. Others limit the term to composite statistics, such as a pupil:teacher ratio, so that the number of students enrolled in a particular phase of education would not be considered an indicator (although it could well be an important item of management information).

A somewhat broader definition is adopted by Shavelson *et al.* (1987): "An indicator is an individual or a composite statistic that relates to a basic construct in education and is useful in a policy context." Yet the authors deny that all statistics are indicators: "Statistics qualify as indicators only if they serve as yardsticks [of the quality of education]." (p. 5)

The confusion over definition was noted by Jaeger (1978), who proposed that:

> "... all variables that *a)* represent the aggregate status or change in status of any group of persons, objects, institutions, or elements under study, and that *b)* are essential to a report of status or change of status of the entities under study, should be termed indicators ... I would not require that reports of status or change in status be in quantitative form, for narrative is often a better aid to comprehension and understanding of phenomena than is a numeric report." (pp. 285-287)

It therefore seems that there is no clear agreement on exactly what an indicator is or is not. Selden (1992) cuts the Gordian knot by proposing that preconceptions about what indicators are should be dropped, as it is their *use* that makes them "indicators".

In this chapter, an indicator is taken to be *quantitative* and could even include the quantified subjective judgement of a professional (*e.g.* the rating of the quality of teaching); *it could also be cited alongside other similar indicators to allow comparison* (usually over time, but also with an average or norm, or with values from other institutions, regions or nations). Above all, indicators are *part of a set or system of indicators* that, together, provide more information than does the sum of its parts; they are not isolated (as test scores have been in some international comparisons of achievement in the past). The idea of an indicator system is discussed below.

Indicators in the policy-making process

If there is no agreement on the definition of indicators, there is a large measure of agreement on their purpose: they are designed to give information to policy-makers about the state of the education system, either to demonstrate its accountability or, more commonly, to help in policy analysis, policy evaluation and policy formulation. The policy-makers can be at the national, regional or district level, within the institution itself, or even in the classroom where, in effect, the teacher reacts to information about the pupils' progress to adjust the pacing or focus of the teaching.

Indicators will naturally be only one of the aids to policy analysis, alongside such techniques as cost-benefit analysis and futures research, but they are nevertheless seen as an increasingly important contribution to rational policy analysis (Carley, 1980; Hogwood and Gunn, 1984). Moreover, indicators tend to send signals about what is or should be important, and thus contribute to public identification of policy issues and concerns – the stream of public problems seen as important, as Kingdon (1984) put it. Indeed, Innes (1990*a*) argued that "social indicators ultimately have their most important role to play in framing the terms of policy discourse" (p. 431). Elsewhere, she proposed an interpretative or phenomenological view of knowledge to help the public understand indicator concepts (Innes, 1990*b*).

Other authors also take the view that research knowledge is not used directly by policy-makers. This is partly because there are limits to the rationality of the policy-making process, as Cohen and Spillane argue (Chapter 16 in this volume), and partly because knowledge is only one element of policy-making, which is inevitably a political process (McDonnell, 1989). Indeed, the only function of knowledge in the policy-making process may be to alter the general climate of opinion (Nisbet and Broadfoot, 1980). Weiss (1979) sees its function as general "enlightenment":

> "Here it is not the findings of a single study nor even of related studies that directly affect policy. Rather it is the concepts and theoretical perspectives that social science research has engendered that permeate the policy-making process." (p. 429; cited by McDonnell, 1989, p. 244)

In the United States, social indicators came to prominence in the 1960s and 1970s but faded from view in the 1980s, and several factors contributed to their decline (Rockwell, 1989; Rose, 1990). The first was essentially political. Indicator systems embody value judgements about what is meant by quality or desirable outcomes in education, and the underlying models or frameworks are not objective. Van Herpen (1992) demonstrates that they almost always are biased towards a particular epistemological model of the education system (for example, the economic or the sociological). The meaning of the indicators (and changes in them over time) thus becomes contentious, and there is a "tendency for indicators to become vindicators" (Bulmer, 1990, p. 410) and for the reports that present them to be "rather bland compromises, deliberately presented without text that might link the data to policy." (Innes, 1990*a*, p. 430)

Secondly, the system became divorced from the policy context and too theoretical and abstruse, run essentially for and by the social sciences community. Innes (1990*a*) suggested that social scientists had an overly simplistic and optimistic view of how and in what circumstances knowledge is used in policy analysis and how straightforward it would be to develop indicators:

> "They focused energy on the measurement task, often to the exclusion of the political and institutional one. They did not recognise how the political and institutional issues would interact with decisions about methodology." (p. 431)

Bulmer (1990) echoed this analysis, and also attributed the lack of success of the social indicators movement to the failure of social science to become institutionalised by the governments of the OECD countries, something that is particularly difficult to achieve under conservative administrations. MacRae (1990) concurred, suggesting a need for a "technical community":

> "... an expert group that conducts and monitors research, but directs its work at concerns of citizens and public officials, not merely at improving its own theories [in the manner of a scientific community]." (p. 437)

The third and, according to Bulmer (1990), the most important factor behind the relative failure of social indicators was the lack of general social science theories specific enough to allow the development of indicators to measure the theoretical constructs. Economic theories have been worked out in much more detail, and economic indicators have the advantage of a common unit of value (money), even though they sometimes include other kinds of numbers such as the unemployment rate. Yet there are rival economic theories and much dispute over the interpretation and explanation of indicators. In other social sciences, "the absence of theory does not preclude the construction of indicators, but it means that ... they often lack a clear rationale and conceptual justification." (p. 409)

How are indicators chosen?

The recent past clearly has much to teach about the factors that ought to be taken into consideration in creating an indicator system in education. Three basic sources of influence appear to interact: policy considerations, scientific/technical issues, and practical issues. These are considered in turn below.

Policy considerations

For information of general interest about the state of the education system, the indicators chosen may simply use information already available. This seems to have been the solution adopted for the Wall Chart in the United States, which displayed only data that were routinely collected for a variety of purposes; the indicators were chosen for their perceived relevance to the educational performance of the 50 states. Some of the indicators have been criticised on the grounds that they do not permit fair comparison. For example, average SAT (Scholastic Aptitude Test) scores need to be corrected for the different proportions (largely due to self-selection) of the student population that take the test in each state, if the comparison is to be meaningful (Wainer, 1986). Socio-economic differences between state populations also ought to be taken into account if the comparisons are to be fair, as has been done in school or school district comparisons in some states (Salganik, 1990; and in school comparisons in the Inner London Education Authority in the United Kingdom (ILEA, 1990). The publication of the Wall Chart has stimulated a number of activities designed to improve the set of indicators displayed, and for that reason alone may be considered to have been a valuable impetus to improvement.

The Panel on Indicators established by the National Center for Educational Statistics in the United States is following a more systematic strategy and is likely to recommend a thematic approach, perhaps with different periodicities for updating. Possible themes include: the acquisition of knowledge and the engagement of the student in the learning process, readiness for entry into school, and equity. The set of indicators for each theme may be arranged in a pyramid fashion, with a few key indicators at the top and many more in tiers below to allow for a more detailed analysis (see Bryk and Hermanson, Chapter 2 in this volume).

Methods that limit the number of indicators tend to give information about current policy issues or about the attainment of particular goals. The targets set by the President of the United States jointly with the state governors in 1990 lend themselves to the creation of particular indicators. The proponents of institutional development planning, who see indicators as the primary tool for evaluating the degree to which the particular targets chosen for a given development cycle are attained, also advocate such an approach (*e.g.* Osborne, 1990; Hargreaves *et al.*, 1989).

A system of indicators based on the policy concerns of the day runs the risk, as Carley (1981) put it, "of faddism, and over-concentration on social factors of passing interest at the expense of those not currently subject to influence and debate" (p. 126). Hogwood and Gunn (1984) took a similar view and advocated including indicators which may not currently appear important or subject to much change over time but might suddenly become significant in the future. Darling-Hammond (Chapter 18 in this volume) also stresses the importance of creating indicators that are independent of the current policy agenda.

While the desire of busy policy-makers and managers for a limited and simple set of indicators, and of researchers for a parsimonious one, is understandable, there are dangers in keeping the set small. The greatest danger is corruption of the behaviour of those whose performance is being monitored. The best-known example is "teaching to the test", most commonly seen when the stakes are high, *i.e.* when the future career of individuals hinges on the results. More generally, education suffers when almost all effort is devoted to changing indicator values for the better. Darling-Hammond therefore argues for a measure of redundancy in the information conveyed by an indicator set, so that behaviour changes in respect to one indicator will also affect others.

McEwan and Hau Chow (1991) take this principle further and argue (p. 81) for a strategy that incorporates what might be called the multiplier effect:

- multiple goals of education, based on appropriate dimensions and domains of schooling;
- multiple indicators of each goal measured by multiple methods;
- multiple levels of analysis: student, class, school, system, province, and, potentially, the country, the world; and
- multiple participants: government, administrators, teachers, academics, parents.

While it is easy to see the value of such a set of indicators, factors such as feasibility and cost curtail the possibility of developing it. The foregoing views emphasize different kinds of issues, yet policy considerations – indeed, the whole policy context – will always remain essential to the usefulness of the indicator system. As McDonnell (1989) sees it:

> "The policy context, then, plays two distinct roles in the design of an indicator system. First, it provides the major *rationale* for developing and operating such a system... Second, the policy context constitutes a key *component* of any educational indicator system, because specific policies can change the major domains of schooling in ways that affect educational outcomes." (pp. 241-242)

The modelling approach

The value of a system of indicators that reveals cause and effect relationships, and that can therefore predict changes as a result of policy-makers' actions, is obvious. However, such a system is not easy to develop. The model must include variables that are amenable to direct manipulation by the policy-maker and that can, through some causal mechanism, affect the desired outcomes.

If social and educational research has, over the years, provided much evidence of relationships between variables, there is currently no general model of the educational process that includes all phases from pre-school to recurrent education and all kinds of outcomes. Van Herpen (1992) shows how many different models have been put forward in educational research and how incomplete and epistemologically biased they are. The OECD Indicators Project adopted a broad framework based essentially on an input-output model of education, and researchers were quick to draw attention to competing models from which different indicator sets can be derived (*e.g.* Fasano, Chapter 3; Bryk and Hermanson, Chapter 2; and Hopkins, Chapter 8, all in this volume). Moreover, Bulmer (1990) claims that theories and models of social indicators are too general to provide an adequate starting point for the development of indicators. Alternatively:

> "Perhaps the sticky wickets are the policies that defy modelling with social indicators, or, if modelled, do not yield answers legislators or administrators want." (Ferriss, 1990, p. 415)

Only econometric models are detailed enough to be used to predict the future behaviour of the economy, but even these models provide various results and reflect the theoretical positions of the modellers. Policy-makers' expectations for such models may be too high. As Greenberger *et al.* (1976), reviewing the use of models in the policy process, put it:

> "... the effectiveness of policy modelling depends not only on the model and the modeler, but on the policy-maker too. Increasing the usefulness of models as instrument for enlightening decision-makers will require behavioural adjustments by the policy-makers as well as by the modelers." (pp. 328-329)

If adequate models cannot be constructed, indicator systems still need some organising principles. The term "framework" is commonly used to avoid implications of cause and effect. Social indicators, for example, often use the structures of health or education programmes. Carley (1980) views this approach as cost-effective and straightforward, but adds:

> "The chief danger is that the sometimes tenuous cause and effect relationships implicit in the indicators might go unnoticed by administrators who may overvalue the explanatory power of the indicators." (p. 194)

The framework put forward by the Rand study on indicators for mathematics and science (Shavelson *et al.*, 1987) presents such a danger. It was constructed after study of the research literature and appears to be a flow chart or model (Figure 4.1).

In the text that accompanies the figure, the authors make a very important caveat, which might easily be missed:

> "The relationships depicted in this figure, of course, do not constitute a model in either a strict predictive or causal sense. However, they can serve as a framework, showing logical linkages among elements of the schooling system." (Shavelson *et al.*, 1987, pp. 10-11)

The general consensus is that present understanding of the educational process is insufficient for the postulation of a model, but that it is possible to create a framework that embodies the available limited knowledge of empirical relationships and that begins to relate malleable variables to desirable outcomes *without* appearing to promise too much. The INES Project has moved cautiously in developing a framework, for this reason among others. It employed a very basic framework in its first phase (Figure 4.2).

In its second phase, the framework has been considerably elaborated, but in each case, arrows between the boxes, which might imply causal relationships, have been avoided.

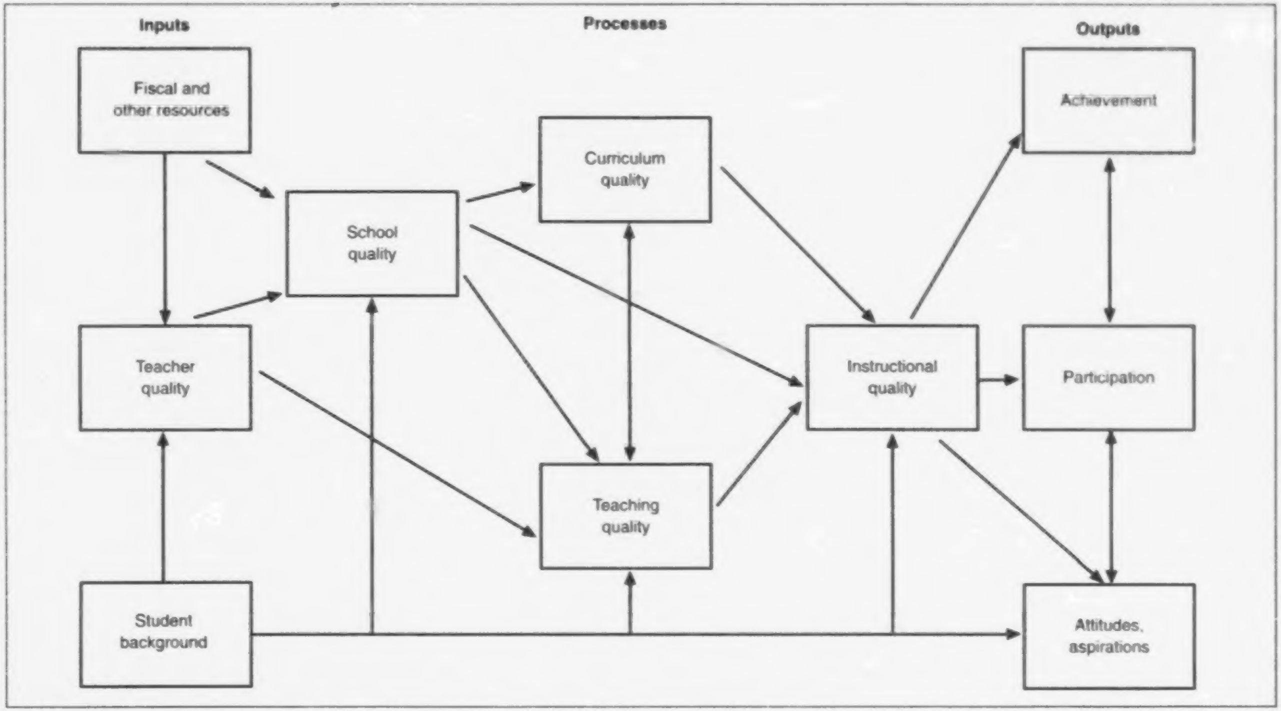
Thus, the two approaches (the one derived from policy considerations and the other from the modelling of the educational process) can be united in the form of a framework, as long as no strong cause and effect relationships are implied and as long as it is recognised that both political and epistemological values have influenced both the general design of the framework and the particular indicator categories used.

Scientific and technical issues

If it is difficult to arrive at a general framework or model that embraces policy-relevant concepts such as "achievement in science" and "quality of teaching", it is also difficult to define concepts precisely enough to allow measures to be taken. Problems involved in moving from concept to measure are well known in the social sciences. One concept can generate dozens of different indicators, and most concepts require detailed specification and clarification. For example, what sorts of skills (in what mix), applied to what facts and concepts, constitute "achievement in science"? Value systems inevitably influence the choice. As a result, it may prove much easier to develop measures of some skills than of others, and it may be much easier to specify certain skills rather than others (Shavelson, 1992). Thus, practical issues such as cost begin to assume importance.

Given the potential subjectivity involved in developing indicators, it is important to evaluate them according to principles of social sciences, so as to judge the reliability and validity of the measures used. Well-established techniques for doing this exist (Messick, 1989; Feldt and Brennan, 1989). For validity, two questions are relevant: How do the measures relate to the concept? How does the framework linking the concepts relate to the reality of the education system? The questions are important because the indicator set may not be an adequate representation of reality. The risk is particularly great for international indicators, notably measurements of student achievement. It is almost impossible to devise a single test that is an equally valid measure of different countries' definition of, for example, achievement in mathematics or science.

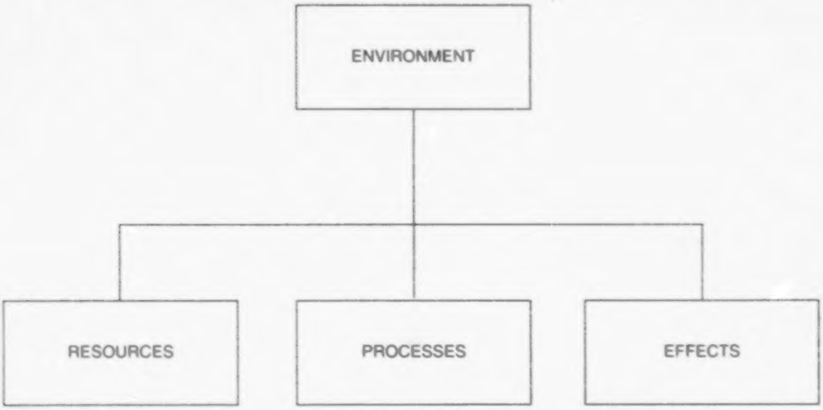
Figure 4.1 **Linking elements of the education system**


Inputs
Processes
Outputs
Fiscal and other resources
Curriculum quality
Achievement
School quality
Teacher quality
Instructional quality
Participation
Teaching quality
Student background
Attitudes, aspirations

Source: Shavelson *et al.* (1987), p. 10.

BEST COPY AVAILABLE

Figure 4.2 **General framework for the definition of international indicators of education systems**



Source: *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, 1992, p. 19.

In practice, the development of indicators and the study of their interdependence can help to refine the concepts and the relationships among them represented in the framework. The interaction of theory and measurement has been particularly productive in the field of school effectiveness (see Scheerens, 1990) and frequently leads to the development of composite or complex concepts. For example, Oakes' review of the literature on the influence of the school context (1989) led her to propose three general constructs that facilitate rather than instigate student learning: access to knowledge, incitement to achieve, and professional teaching conditions. All three are far from simple, and for each Oakes suggests that it would be necessary to measure at least nine "more tangible school characteristics". She does not propose whether or how the separate measurements should be combined to form a single indicator, or how their validity would be established.

The creation of composite indicators is likely to be important in indicator systems, if only to avoid overloading the reader with numbers. Social sciences and statistics can again offer well-tried techniques for forming homogeneous composite measures – for example, using factor analysis or other approaches to multidimensional scaling (see Mardia *et al.*, 1979) – but they have rarely been applied in indicator systems (except in achievement and attitude measurement). Some authors have gone so far as to suggest a single composite indicator of the success of the education system, namely the Gross Educational Product (analogous to the Gross Domestic Product which is itself composed of a multitude of smaller measures). The former is not, of course, as simple as the latter, since education lacks the sort of common measure that economics employs (dollars or pounds sterling). The creation of a composite would therefore require the application of scaling and weighting techniques (Petersen *et al.*, 1989). In any case, even econometric modelling runs into difficulties when there is no direct and simple way of generating a

BEST COPY AVAILABLE

monetary indicator, *e.g.* for intangibles such as environmental pollution or for non-monetised activities such as household chores.

Nevertheless, the general thrust of research in education suggests that complex concepts of the kind presented by Oakes are likely to have explanatory powers. Many of these may have to be appraised and measured by experts such as inspectors who may be able to arrive at numerical judgements of the relative "quality of the learning environment" or "professional teaching conditions" across different institutions. Here, not only is validity a concern, but the reliability or consistency of judgement, between experts and over time, also becomes of great significance. Even the reliability of relatively "simple" indicators such as pupil:teacher ratio is not easily achieved.

Practical considerations

To be useful to the policy-maker, indicators not only have to be relevant but, experience shows, they must also be timely, comprehensible and few in number. Ensuring timeliness puts pressure on the indicator technicians, who themselves may be dependent on data provided by thousands of others who may attach little priority to the activity, not least because they often stand to gain little from it. It also affects the cost: modern information technology and sampling offer ways of streamlining procedures, but the cost of providing timely information of high quality is likely to remain substantial. This is particularly true for student achievement measures.

A restricted number of indicators may gain attention and be more readily comprehensible, but it may provide a less valid set of indicators as a framework that represents the education system. Clear presentation will aid comprehension, but many (*e.g.* Odden, 1990) argue that information, however clearly presented, still needs interpretative comment, as indicators do not "speak for themselves". Comment almost inevitably reflects the values of the commentator and thus can lead to political difficulties. When social indicators fell into this trap in the United States, discussions were presented elsewhere (notably in the *Annals of the American Academy of Political and Social Sciences*) and were therefore infrequently referred to.

Another desirable characteristic of indicators is that the data should be incorruptible, in other words not liable to deliberate alteration before they are collected. Achievement testing is particularly prone to such manipulation; schools may ensure that those of lower achievement are absent on the day of the test or may coach students for the test.

Who chooses the indicators?

The previous section has examined the various factors that might influence the way in which indicators and indicator systems are chosen. These choices are inevitably influenced by the value systems of those making the choice and need to reflect the interests of the policy-makers. They must also reflect scientific understanding of how the education system functions. This led MacRae (1985) to propose a "technical community" that could bridge the gap between the policy-makers and the social scientists. Other authors feel that the "consumers" of education should also have a voice (see Riley, 1990). Pollitt (1986) warns that imposing a system of indicators from above will alienate those whose assistance and goodwill are needed in the enterprise, and advocates a pluralistic stance so that every interested party can contribute to the discussion before the system is finalised.

Thus, the answer to the question, "Who chooses the indicators?", must be in large measure political and will inevitably be of significance to the selection of indicators and to the outcomes of that process. Many writers have attempted to recognise its importance alongside political, technical and practical considerations, by creating a set of criteria that an indicator system, and individual indicators, should meet before they come into use.

Criteria for choosing, developing and evaluating indicators

These criteria differ according to political values and according to the particular policy context, education system and decision-making level under study. The principles that have been proposed for developing and evaluating indicators therefore vary, as do the safeguards that have been advocated against misinterpretation and misuse, but the criteria proposed also have much in common.

For the OECD Project on International Education Indicators, Nuttall (1992, p. 19) proposed the following six principles:

- Indicators are diagnostic and suggestive of alternative actions, rather than judgemental.
- The implicit model underlying a set of indicators must be made explicit and acknowledged.
- The criteria for the selection of indicators must be made clear and related to the underlying model.
- Individual indicators should be valid, reliable and useful.
- Comparisons must be done fairly and in a variety of different ways...
- The various consumers of information must be educated about its use.

Except for the more technical criteria, this list appears to be primarily concerned with lowering the expectations and increasing the sophistication of users.

A further criterion, aimed at safeguarding against the punitive use of indicators, was originally proposed: control over data must remain with those who provide it. As desirable as such a criterion might be, it was not retained on the grounds that in an international project it could not be met.

Other criteria proposed by English writers have attempted to offer a pluralistic view of indicators so as to ensure that the community has a genuine stake in them. For example, Riley (1990) proposed the following set:

- The process of developing school indicators should ensure that all the partners in education have a sense of ownership in the indicators.
- The indicators should be accessible to all the partners in education.
- They should be comparable throughout the levels of authority in the system, even at school district or local education authority level.
- They must be linked to school ethos and objectives.
- They should be inclusive of both cognitive and non-cognitive outcomes.
- The indicators should be implementable.
- They should be based on consumer evaluation of the education experience.

It is apparent that this set concentrates on school-level indicators, as does the set proposed by Gray (1990):

- The most important consideration relating to the construction of performance indicators is that they should directly measure or assess schools' performance.
- They should be central to the processes of teaching and learning which are schools' prize objectives.
- They should cover significant parts of schools' activities but not necessarily (and certainly not to begin with) all or even most of them.
- They should be chosen to reflect the existence of competing educational priorities; a school which did well in terms of one of them would not *necessarily* be expected (or found) to do well in terms of the others.
- They should be capable of being assessed: assessment is distinguished here from measurement, which implies a greater degree of precision than is intended.
- They should allow meaningful comparisons to be made over time and between schools.
- They should be couched in terms that allow schools, by dint of their efforts and the ways in which they choose to organise themselves, to be seen to have changed their levels of performance; that is, to have improved or, alternatively, to have deteriorated relative to previous performance and other schools.
- They should be few in number: three or four might be enough to begin with. After some experimentation over a period of years one might end up with a few more.

Gray (1990) went on to propose what those three indicators should in fact be, based on the research evidence from the school-effectiveness literature. His proposal is shown in Table 4.1.

The writer sees particular merit in keeping the number of indicators small and is not concerned that the list only partially covers the goals of education. The indicators focus on important goals and pay no attention to processes and virtually none to context (only questions 1a and 1b, by using the term "expected levels of progress", acknowledge the relativity of the measures).

Criteria proposed in the United States tend to reflect primarily the concerns of policy-makers above the level of the school. Windham (1990) drew attention to the conclusion that economic indicators should be "accurate, relevant, timely, understandable and affordable". Carley (1981) attributed the failure of the social indicators movement largely to the influence of researchers who, in the search for accuracy, ignored relevance, timeliness and comprehensibility. Nevertheless, the importance of these factors had been recognised for some time. For example, the U.S. Urban Institute put forward the following criteria (Hatry, 1977, quoted by Carley, 1981, p. 166):

- *Appropriateness and validity.* Indicators must be quantifiable, in line with goals and objectives for that service, and be oriented towards the meeting of citizen needs and minimising detrimental effects.
- *Uniqueness, accuracy and reliability.* Indicators generally need not overlap, double counting should be avoided, but some redundancy may be useful for testing the measures themselves.
- *Completeness and comprehensibility.* Any list of indicators should cover the desired objectives and be understandable.
- *Controllability.* The conditions measured must be at least partially under government control.

Table 4.1 **Three indicators derived from the research on school effectiveness**

Indicator focus		Key questions	Response categories
1. *Student progress*			
	1*a*)	Taking the school as a whole, what proportions of students made expected levels of progress over the relevant time period?*	All or most Well over half About half Well under half Few
	1*b*)	What proportion of students of: *i*) below average *ii*) average *iii*) above average prior attainment made expected levels of progress over the relevant time period?*	All or most Well over half About half Well under half Few
2. *Student satisfaction*			
	2*a*)	What proportion of students in the school are satisfied with the education they are receiving?	All or most Well over half About half Well under half Few
	2*b*)	What proportion of students of: *i*) below average *ii*) average *iii*) above average attainment are satisfied with the education they are receiving?	All or most Well over half About half Well under half Few
3. *Student-teacher relationship*			
	3*a*)	What proportion of students in the school have a good relationship with one or more teachers?	All or most Well over half About half Well under half Few
	3*b*)	What proportion of students of: *i*) below average *ii*) average *iii*) above average attainment in the school have a good relationship with one or more teachers?	All or most Well over half About half Well under half Few

* Initially, this question might be posed in terms of summary measures of pupil attainment; subsequently, more detailed breakdowns (subject by subject, for example) might be attempted.

- *Cost.* Staff and data collection costs must be reasonable.
- *Feedback time.* Information should become available within the time frame necessary for decision-making.

Rockwell (1989) offers a rather similar list: timeliness; providing "handles for policy"; covering both current and emerging policy issues; in a time series, measures adaptable to changing circumstances – valid, reliable and accurate.

The most influential proposals for indicator criteria in the United States emerged from the Rand work on systems for monitoring mathematics and science education (Shavelson *et al.*, 1987). They were that indicators should:

- *Provide information that describes central features of the educational system, e.g.* the amount of financial resources available, teachers' work load, and school curriculum offerings. Even though research has not as yet determined the relationship of some of these features to particular outcomes, information is needed about them to understand how the system works and because policy-makers and the general public care about factors such as per pupil expenditures and class size.
- Provide information that is *problem-oriented.* Indicators must provide information about current or potential problems – for example, factors linked to teacher supply and demand, or to the changing demographics of urban areas.
- Provide information that is *policy-relevant.* Indicators should describe educational conditions of particular concern to policy-makers and amenable to change by policy design. For example, indicators of teacher characteristics such as educational background and training are relevant to policy, since they can be changed through legislation or regulations governing the licensing of teachers.
- *Measure observed behaviour rather than perceptions.* Indicators will be more credible if they assess actual behaviour rather than participants' opinions or judgements. For example, the academic rigour of schools is better measured by course requirements and offerings than by the perceptions of the principal, teacher and student.
- *Provide analytical links among important components.* Indicators will be more useful if they permit the relationships among the different domains of schooling to be explored.
- Generate data from measures that are generally accepted as *valid and reliable.* Indicators should measure what they are intended to measure and should do so consistently.
- Provide information that can be *readily understood by a broad audience.* Indicators need to be easily comprehensible and meaningful to those beyond the immediate mathematics and science community – to policy-makers, the press and the general public.
- Be *feasible* in terms of timeliness, cost and expertise. Indicator data need to be produced within a time frame that is compatible with policy-makers' decision cycles and within given cost constraints: they should also be collectable, analysable and reportable with current levels of expertise.

On the more technical side, this list is similar to the others (several of which it influenced, no doubt) but it again sees the policy-maker as the main client for indicator information.

Some of the differences between the lists are a function of the particular audience that the indicators are designed to address (*e.g.* national policy-makers or school personnel) but other differences are less easy to explain. There is broad agreement about technical and practical matters such as validity, reliability, timeliness, comparability, feasibility and keeping costs reasonable, and little difference on the need for policy relevance and the importance of ensuring that the indicators are comprehensible to their audiences.

The major areas of difference are the number and focus of the indicators. While most authors recognise that the indicators (or at least some of them) should be alterable or controllable (by the actions of the policy-makers), they differ about their number, the need for redundancy, and the extent to which the indicators should be comprehensive and organised by and into a framework that reflects the functioning of the education system, with, where possible, known causal links. Where the set is not comprehensive, there is agreement that it should focus on the central features and outcomes of the educational process.

It is therefore unavoidable that indicators cannot meet the different criteria that have been proposed. The developers of an indicator set must decide whether they are going to lean more towards a small number of key indicators or more towards a comprehensive set, embodying context and process as well as outcome. They will have to trade one criterion off against another, *e.g.* greater comprehensiveness against greater cost.

Conclusion

This chapter has attempted to describe and analyse the cluster of interacting factors that influence the development of an indicator system. These factors are: policy considerations; research knowledge; technical considerations; practical considerations; and the "choosers", those in a position to influence the choice and development of indicators.

Many authors have attempted to indicate, by stating criteria, how these factors can be translated into a set of principles for guiding the development of indicators. The previous section has shown that indicators should be:

- policy-relevant;
- policy-friendly (timely, comprehensible and few in number);
- derived from a framework (defensible in research terms and including alterable variables, and thus oriented towards action);
- technically sound (valid and reliable);
- feasible to measure at reasonable cost.

The nature and importance of these considerations will vary according to the locus or level of the action and the purpose of the system of indicators. For example, the framework and the potential action points for an indicator system led primarily by the concerns of national policy-makers might be different from those for one designed for a local school system or an individual school site. The differences might be even more pronounced between a system designed to inform managers at the local level and a system designed for local accountability, which would tend to stress outcomes much more.

It also must be recognised that, whatever the level, these principles interact and sometimes conflict. Increases in validity rarely occur without increases in cost, and may well adversely affect timeliness. As a result, the quality of the indicator system is likely to be determined by those who hold the purse-strings, and they are almost always the policy-makers. It follows, first, that policy relevance and policy friendliness are likely to be of major significance, possibly at the expense of the "scientific" validity of the framework. Second, fewer rather than more indicators are likely to be preferred, again possibly at the expense of validity. Finally, the technicians will have to work closely with the policy-makers to ensure that expectations do not run so high that they are fated to become disillusionments, while still pointing to the potential, albeit limited, value of developing an indicator system.

Bibliography

BULMER, M. (1990), "Problems of Theory and Measurement", *Journal of Public Policy*, Vol. 9 (4), pp. 407-412.

CARLEY, M. (1980), *Rational Techniques in Policy Analyses*, Heinemann Educational Books, London.

CARLEY, M. (1981), *Social Measurement and Social Indicators: Issues of Policy and Theory*, George Allen & Unwin, London.

CIPFA (1988), *Performance Indicators in Schools: A Contribution to the Debate*, CIPFA (The Chartered Institute of Public Finance and Accountancy), London.

FELDT, L.S. and BRENNAN, R.L. (1989), "Reliability", in Linn, R.L. (ed.), *Educational Measurement*, 3rd edn., American Council of Education and Macmillan Publishing Company, New York.

FERRISS, A.L. (1990), "Whatever Happened, Indeed!", *Journal of Public Policy*, Vol. 9 (4), pp. 413-417.

GRAY, J. (1990), "The Quality of Schooling: Frameworks for Judgement", *British Journal of Educational Studies*, Vol. 38 (3), pp. 204-223.

GREENBERGER, M., CRENSON, M.A. and CRISSEY, B.L. (1976), *Models in the Policy Process*, Russell Sage Foundation, New York.

HARGREAVES, D.H., HOPKINS, D., LEASK, M., CONNOLLY, J. and ROBINSON, P. (1989), *Planning for School Development: Advice to Governors, Headteachers and Teachers*, Department of Education and Science, London.

HATRY, H.P. (1977), *How Effective Are your Community Services: Procedures for Monitoring the Effectiveness of Municipal Services*, The Urban Institute, Washington, D.C.

HOGWOOD, B.W. and GUNN, L.A. (1984), *Policy Analyses for the Real World*, Oxford University Press, Oxford.

ILEA (Inner London Education Authority) (1990), *Differences in Examination Performance* (RS 1277/90), ILEA Research and Statistics Services, London.

INNES, J.E. (1990*a*), "Disappointments and Legacies of Social Indicators", *Journal of Public Policy*, Vol. 9 (4), pp. 429-432.

INNES, J.E. (1990*b*), *Social Indicators and Public Policy: Interactive Processes of Design and Use* (2nd edition), Transaction Books, New Brunswick, NJ.

JAEGER, R. (1978), "About Educational Indicators: Statistics on the Conditions and Trends in Education", *Review of Research in Education*, Vol. 6, pp. 276-315.

JOHNSTONE, J.N. (1981), *Indicators of Education Systems*, Kogan Page and UNESCO, London and Paris.

KINGDON, J.W. (1984), *Agendas, Alternatives, and Public Policies*, Little, Brown and Company, Boston, MA.

MacRAE, D. Jr. (1985), *Policy Indicators: Links between Social Science and Public Debate*. The University of North Carolina Press, Chapel Hill, NC.

MacRAE, D. Jr. (1990), "Policy Indicators: A Continuing and Needed Field", *Journal of Public Policy*, Vol. 9 (4), pp. 437-438.

MARDIA, K.V., KENT, J.T. and BIBBY, J.M. (1979), *Multivariate Analyses*, Academic Press, London.

McDONNELL, L.M. (1989), "The Policy Context", in Shavelson, R.J., McDonnell, L.M. and Oakes, J. (eds.), *Indicators for Monitoring Mathematics and Science Education: A Sourcebook*, Rand Corporation, Santa Monica, CA.

McEWAN, H. and HAU CHOW (1991), "Issues in Implementing Indicator Systems", *Alberta Journal of Education Research*, Vol. 37 (1), pp. 65-86.

MESSICK, S. (1989), "Validity", in Linn, R.L. (ed.), *Educational Measurement* (3rd edition), American Council of Education and Macmillan Publishing Company, New York.

NISBET, J. and BROADFOOT, P.M. (1980), *The Impact of Research on Policy and Practice in Education*, Aberdeen University Press, Aberdeen.

NUTTALL, D.L. (1992), "The Functions and Limitations of International Education Indicators", in *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, pp. 13-23.

OAKES, J. (1989), "What Educational Indicators? The Case for Assessing School Context", *Educational Evaluation and Policy Analysis*, Vol. 11 (2), pp. 181-199.

ODDEN, A. (1990), "Making Sense of Education Indicators: The Missing Ingredient", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 33-50.

OSBORNE, D.A. (1990), "The NSW School Development and Evaluation Model", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 147-154.

PETERSEN, N.S., KOLEN, M.J. and HOOVER, H.D. (1989), "Scaling, Norming and Equating", in Linn, R.L. (ed.), *Educational Measurement* (3rd edition), American Council of Education and Macmillan Publishing Company, New York.

POLLITT, C. (1986), "Performance Measurement in the Public Services: Some Political Implications", *Parliamentary Affairs*, Vol. 39 (3), pp. 315-329.

RILEY, K. (1990), "Making Indicators Consumer-Friendly", *Education*, 11 May, pp. 470-472.

ROCKWELL, R.C. (1989), "Lessons from the History of the Social Indicators Movement". Prepared for the National Center for Education Statistics (NCES) Indicators Panel, Washington, D.C., October.

ROSE, R. (ed.) (1990), "Whatever Happened to Social Indicators? A Symposium", *Journal of Public Policy*, Vol. 9 (4), pp. 399-450.

SALGANIK, L.H. (1990), "Adjusting Educational Outcome Measures for Student Background: Strategies Used by States and a National Example". Prepared for the National Center of Education Statistics (NCES) Indicators Panel, Washington, D.C., October.

SCHEERENS, J. (1990), "School Effectiveness Research and the Development of Process Indicators of School Functioning", *School Effectiveness and School Improvement*, Vol. 1 (1), pp. 61-80.

SELDEN, R. (1991), "The INES Framework", in *The INES Handbook*, OECD/CERI, Paris (unpublished manuscript).

SELDEN, R. (1992), "Standardizing Data in Decentralised Educational Data Systems", in *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, pp. 107-113.

SHAVELSON, R.J. (1992), "Can Indicator Systems Improve the Effectiveness of Mathematics and Science Education? The Case of the U.S.", *Evaluation and Research in Education.*

SHAVELSON, R.J., McDONNELL, L., OAKES, J., CAREY, N. with PICUS, L. (1987), *Indicator Systems for Monitoring Mathematics and Science Education,* The RAND Corporation, Santa Monica, CA.

van HERPEN, M. (1992), "Conceptual Models in Use for Education Indicators", in *The OECD International Education Indicators: A Framework for Analysis,* OECD/CERI, Paris, pp. 25-51.

WAINER, H. (1986), "The SAT as a Social Indicator: A Pretty Bad Idea", in Wainer, H. (ed.), *Drawing Inferences from Self-Selected Samples,* Springer-Verlag, New York.

WEISS, C.H. (1979), "The Many Meanings of Research Utilisation", *Public Administration Review,* Vol. 39, pp. 426-431.

WINDHAM, D. (1990), Personal Communication (Comments made as Discussant at a Symposium on the INES-OECD/CERI Project at the Annual Meeting of the American Educational Research Association, April).

Part Two

DEVELOPMENT OF INDICATORS

Chapter 5

Education Indicators: A Review of the Literature

by

Tim Wyatt
Washington Consulting Group
Washington, D.C., United States

This chapter examines some of the trends in the now voluminous literature on education indicators. It is neither an exhaustive nor a chronological review but an attempt to chart some of the major issues that have arisen during the past decades. The chapter focuses primarily on school-level indicators. While some concepts apply to all education levels, issues relating specifically to the higher education sector are generally excluded. The literature concerning higher education indicators is itself large but has been well covered by Frackman (1987) and the supplement and addendum to the *Frascati Manual* (OECD, 1989).

With the number of publications increasing almost daily, it is necessary to limit the scope of this review. Two relatively complete bibliographies are available which cover the main conceptual and methodological themes. The first of these, *Educational Quality Indicators*, is an annotated bibliography of 230 publications on indicators drawn up by the Alberta (Canada) Ministry of Education (1990). The second source is a bibliography prepared as part of a project initiated by the Australian Conference of Directors-General of Education (Wyatt *et al.*, 1988).

The organisation of this chapter is as follows: first, the literature on the subject is summarised in order to provide a background to the current situation and the dominant themes. Each theme is then examined in depth, drawing on major reviews to identify trends in theory and practice. The chapter concludes with a discussion of the prospects for future development.

Antecedents of education indicators

The concept of education indicators defined as summary statistics on the current status of education systems is not new. Studies of the effectiveness of schools can be traced almost to the beginning of mass education. These early evaluation studies often took the form of an examination of the "standards" attained by pupils and the associated "payment by results" schemes decried by Matthew Arnold in *The Twice Revisited Code*

BEST COPY AVAILABLE

as early as 1862. Throughout the late 19th and early 20th century, ministers' reports to parliament, inspectors' accounts and reports to district, county and state boards of education were important vehicles for disseminating information about the condition of schooling in many countries.

Madaus (1981) has commented that whenever there are perceptions of falling levels of achievement, the traditional response has been a call for greater accountability and the imposition of higher standards. Shavelson *et al.* (1989) note that the concerns raised and solutions proposed during the Common School Movement in the 1880s, and the rationale for the establishment of the U.S. National Center for Education Statistics, are similar to those expressed today.

Current ideas about performance and education indicators can be seen as an extension of the research conducted on social indicators by social scientists in the mid-1960s. That work had its own origins in the attempts to measure social change by William O'Brien and colleagues at the University of Chicago in the 1930s and 1940s. The movement began with high hopes and ended with a retreat from ambition. Detailed accounts of the development of the social indicators movement and its predecessors have been described by Land (1975), de Neufville (1975), Carley (1981) and Broad (1983). Wyatt (1988*b*) and Rockwell (1990) have reviewed the literature on social indicators and identify a number of lessons that can be drawn from the history of that movement.

Interest in social indicators was partly a reaction to the successful use of economic indicators in the early 1960s to guide government policy (de Neufville, 1975). The inability of economic indicators to provide evaluations of wider social welfare considerations, such as qualitative aspects of life, equity and the side effects of economic prosperity, generated a demand for a systematic framework of social accounting.

Further interest in social indicators in the United States came in 1962 from the American Academy of Arts and Sciences' examination of the second-order effects of the space exploration programme. The research team turned its attention to the general issue of monitoring socio-economic change, the result being the publication of the seminal report, *Social Indicators* (Bauer, 1966). This volume discussed the development of social indicators, their relationship to social goals and policy-making, and the need for systematic social accounts and improved statistical information.

Other significant works followed, including a series of studies on the conceptual and methodological problems of monitoring, such as *Indicators of Social Change: Concepts and Measurements* (Sheldon and Moore, 1968) and *The Human Meaning of Social Change* (Campbell and Converse, 1972), which proposed different approaches to the subject: the first with objective, socio-structural indicators, the second with subjective indicators of attitudes, expectations, aspirations and values.

Other major publications were *Toward a Social Report* by the United States Department of Health, Education and Welfare (1969), which set out the parameters and requirements for the development of a comprehensive social report, and the United Kingdom Central Statistical Office bulletin, *Social Trends* (1970, 1971). In Australia, comprehensive work was undertaken by the Australian Bureau of Statistics, which led in 1976 to the publication of a *Social Indicators* bulletin. Japan, Sweden and the Federal Republic of Germany also produced high-quality social reports.

The interest of government agencies was complemented by the work of international organisations, with the OECD and UNESCO both fostering the development of social indicators. Of particular significance was an OECD study on indicators for measuring the impact of education on society (OECD, 1973). UNESCO, on the other hand, tended to view indicators as essential components of evaluation exercises (Fanchette, 1974; UNESCO, 1980).

Social indicators were applied to many areas in the early 1970s, including education, health, recreation, quality of life, and employment. By the late 1970s, enthusiasm for these indicators had waned. Their early promise had for the most part not been realised. Policy-makers had made little systematic use of indicators, and some administrators even dismissed them as useless or misleading (Carley, 1981). Many researchers turned from a concern with social data for informing government decision-making to a perspective involving complex issues in measurement. For many, social indicators were an idea whose time had come and gone.

Carley (1981) offers three explanations. First, expectations of social indicators were too high regarding both the time needed for their elaboration and the adequacy of social theory. Second, eagerness to supply social indicator data to assist in informing policy decisions often led to the provision of *ad hoc* information of poor quality, thereby undermining confidence in this kind of information. Third, there were insufficient attempts to relate indicators explicitly to policy objectives.

Current thinking on education indicators has also been influenced by the success of economic indicators such as the gross national product, consumer price index, exchange rate, inflation rate, unemployment rate, and so on. These indicators now have a well-established and indispensable role in public and private decision-making. There are many similarities in the historical development of economic and education indicators. Both, for example, grew out of necessity or perceived crisis. The measurement of employment, for instance, became crucial during the great depression of the 1930s, while the need for better measures of student outcome was sparked by fears of apparently falling standards publicised in the United States Wall Chart in the 1980s.

Originally, the primary purpose of economic indicators was to track the magnitude and direction of changes in specific types of economic activities at the national level. However, their scope eventually expanded to include predictive, planning, analytical and evaluative capabilities in measuring progress towards economic targets and objectives (Horn and Winter, 1990).

A number of separate phenomena caused a revival of interest in indicators. It coincided with the scrutiny of education in the United States that was sparked by a report of the National Commission on Excellence in Education, *A Nation at Risk* (1983). This review showed how little was known about schools and schooling in this country. Solid information to substantiate assertions about declining levels of achievement were not – and are still not – collected regularly. Further interest was generated by the publication of the Wall Chart, the purpose of which was to provide comparative data that indicated key features – inputs, processes and outçomes – of each of the 50 U.S. states. The generally negative reception of the chart by the scientific community led to a number of major studies, notably those by the Rand Corporation and the Center for Policy Research in Education at Rutgers University.

A report by Oakes, *Education Indicators: A Guide for Policymakers* (1986), was particularly significant. The definitions and criteria she proposed were helpful in avoiding mistakes previously made in the work on social indicators, and the report has influenced many other authors (see, for example, Fuhrman, 1988).

Further legitimacy for the concept of indicators as a vehicle for accountability and public reporting was added when the National Center for Education Statistics decided to publish its annual report, *The Condition of Education* (1987) in the form of indicators, followed by the publication of *Youth Indicators* (National Center for Education Statistics, 1989).

Also partly in response to the Wall Chart, the Council of Chief State School Officers reversed their stand of 20 years or more to create the State Education Assessment Center to produce new, accurate and valid comparative information on state education systems. The improvement of the National Assessment of Educational Progress (Alexander and James, 1987) has been an important part of that exercise. The Center has, in addition, been at the forefront of research issues related to indicator development and use, including most recently how to present data that appropriately accounts for different socio-demographic contexts. Meanwhile, many states had begun to develop their own indicator schemes (Honig, 1985; Fetler, 1986).

At the international level, publication of the results of the IEA Second International Science Study (Postlethwaite and Wiley, 1992) aroused considerable interest, reviving an examination of many of the issues involved in making cross-national comparisons. The two OECD international conferences, held in Washington in 1987 and Poitiers in 1988, which led to the establishment of the OECD international indicators project, were influential in sustaining and institutionalising this interest. Although the conferences did not lead to agreement on specific indicators or even indicator areas as the organisers had hoped, they did direct attention to important issues emerging at that time and provoked more in-depth investigation as to how a national picture of the status of education could be constructed. These conferences also highlighted the differences between national methods of collecting, processing and using data in the management of education systems. The OECD indicators project itself has provided a stimulus for conceptual development of indicator methodology and thought (Bottani and Delfau, 1990).

In the United States, the most tangible and powerful symbol of recent political interest in education was the presidential summit held in 1989 in Charlottesville, where the President and state governors met together for only the third time in American constitutional history and discussed, for the first time, education (Griffith, 1990). Following on from the summit was the creation of an Indicators Panel, charged with developing a means of reporting indicators to Congress on an annual basis. The work of the panel has generated a large number of papers on many indicator themes but they are not yet widely available.

Trends in the literature

During the 1980s, public education systems in many countries were faced with diminishing budgets and calls for greater accountability of the use and results of resource spending. What was needed was some means that could simply, accurately and in a timely manner answer these calls. Indicators (going by a variety of names) were proposed as one solution. Cuttance (1989) in Scotland, and Ruby *et al.* (1989) in Australia provide comprehensive reviews of the education indicators debate (McCollum and Turnbull, 1990; Nuttall, 1992). Much of the indicators literature addresses this theme of accountability.

The 1980s was also a decade of reform in the governance of education in many countries, leading to the devolution of responsibility to the school. Indicators were proposed as a means of monitoring change. A second major theme to emerge in the indicators literature concerns their use at the school level. Within this literature – emanating most frequently from the United Kingdom, Australia, New Zealand, the Netherlands and France – a number of sub-themes can be identified. The first is essentially a reflection of the desire of central governments to know what is happening in their decentralised system, and a call for schools to demonstrate accountability with respect to a number of (usually) centrally determined criteria. The debate revolves around what the criteria should be, how they should be measured, and how they should be reported (Cuttance, 1989; Nuttall, 1992).

The second sub-theme is concerned with how schools might evaluate themselves, and emphasizes the use of locally determined indicators in the school management process. A third sub-theme examines the use of indicators to monitor specific policy objectives in schools. The monitoring of academic achievement has been considered important in the United States, with curriculum reform (particularly in mathematics and science courses), teacher supply and demand, participation and drop-out rates, and equity constituting additional evaluation criteria.

A number of other themes intersecting those mentioned above include the problems and limitations of indicators and how they might be improved. Each of these themes is examined below, drawing on the major works addressing the issues.

The rationale for using indicators

The question of why indicators have achieved such prominence in a short time span has not often been explicitly addressed. The usual explanation – that it is a combination of demand for information, combined with the development of the technological capability to provide it (Stern, 1986) – has seldom been examined and is taken as an article of faith by many.

However, many have noted that, in the United States at least, there was increasing frustration with the quality of national data on education (Kaagan and Smith, 1985; Smith, 1988; Selden, 1990; Nuttall, 1992). They also note that calls for better information meant not only more data but that which answered other questions – not how big is the education system, but how well is it doing? Kaagan and Coley (1989) note that governors, legislators and state education agency leaders are no longer operating on the edges of school business, but instead are wrestling with matters of school effectiveness, teacher policy, curriculum emphases and student performance. This higher level of involvement itself has created demands for additional and more specific information.

The indicators concept has also attained popularity in countries such as Sweden and the Netherlands, who have had a comprehensive data collection on education for some time. Part of the answer, identified in an intriguing analysis of why so many have been willing to invest in indicator development (Ruby, 1989), derives from the technical and conceptual criteria required of indicators as now defined. Not only should indicators provide policy-relevant and problem-oriented information, but also information that can be understood by different audiences, and that can be delivered in a manner timely enough to influence decision-making. Ruby notes that there is a need for a reference point in the decision-making process and in the minds of those making policy decisions. Policy-makers generally need information to define the competences and skills of the workforce to track and monitor reforms; to support ideological commitments; and to effect qualitative changes in teaching and learning.

Informed decisions obviously require information that is timely, accurate and accessible. Whether indicator systems prove to do so in the long term remains to be seen. Few attempts to evaluate the effectiveness of indicator schemes exist, yet, except in the form of reports of changes in methodology. Although many point to the potential problems besetting indicators, few challenge the conceptual need for them. Eide (1987) notes that most decisions are not made on the basis of statistical information. Qualitative information has always played a major role in decision making; however, whether this renders indicators useless remains to be seen.

What are education indicators?

Definitions of indicators vary widely, as do the names by which they are known – performance indicators, education indicators, education performance indicators, quality indicators, workload indicators, management indicators, indicators of success (Ashenden, 1987*a*). There appears to be little consistency in how these terms are defined and applied – with the exception of Ruby and Wyatt (1988), who prefer the term "education indicator" to emphasize that the concern is not only with outcomes but the entire education process. Cuttance (1989), in contrast, argues that since education is multilevel, with input at one level being output from another, and because performance may be viewed as a process, the term "performance indicator" would be appropriate.

Definitions are of course important since they determine what can be measured and why, but it is often difficult to interpret differences between definitions. Jaeger (1978), in a review of a dozen definitions, found much that was contradictory and little that was concise or illuminating. To an extent, this is still true today, but perhaps for different reasons.

This lack of consensus comes in part from the influence of the social indicator movement, where there was also little agreement as to definition, scope and purpose. This is illustrated by the variance between major approaches, such as those of the OECD and the United Nations, and in the syntheses of Land (1975) and Jaeger (1978) who found only three common elements: that indicators should be quantitative, measure social conditions, and be time-related or presented in time series. Jaeger (1978, p. 276) recommended that:

> "... all variables that 1) represent the aggregate status or change in status of any group of persons, objects, institutions or elements under study, and that 2) are essential to a report of status or change of status of the entities under study or to an understanding of the condition of the entities under study, should be termed indicators. I would not require that reports of status or change in status be in quantitative form, for narrative is often a better aid to comprehension and understanding of phenomena than is a numeric report."

His recommendation that the definition of an indicator be left open and determined pragmatically appears to have gained some following, although many people now refer to "indicators" either explicitly or implicitly as "statistics".

Stern and Hall (1988, p. 1), for example, regard indicators as:

> "... usually derived statistics which may be either single statistics, such as average teacher salary and average student achievement scores on standardised tests, or composite statistics formed by combining two or more related variables."

More complex measures can be created in this way, in the form of ratios or indices. Some authors distinguish between indicators and variables, which are seen only as the building blocks from which true (*i.e.* composite) indicators are built. Current reviews (for example, McCollum and Turnbull, 1990) tend to present definitions based on amalgamations of previous work. Some, however, refer to indicators only as "pieces of information" (Ashenden, 1987*c*).

An emerging view, based on the new ideas of what the purpose of an indicator should be, sees indicators as statistics that report the "health" of the education system – what students know and can do, the quality of the system's operations, and whether these conditions are improving or declining over time (Smith, 1984). A distinction is drawn between the more common education measures of size or level (*e.g.* numbers of students, schools or teachers) which, while offering interesting facts about education, do not directly address the question of "health".

More explicitly, others propose that indicators are intended to aid in policy analysis. This is achieved by developing indicators that are capable of being contrasted with a standard criterion level or with themselves over time, compared across systems, and contrasted with other indicators in a cost-benefit model (Kaagan and Smith, 1985). Whether this latter need for a reference point is part of the statistical calculation formula for the indicator, or whether it merely refers to how the information is presented, is not always clear.

The most frequently cited definition in recent times derives from the composite definition of Oakes (1986, pp. 1-2) who argues that indicators must provide at least one of the following types of information:

a) information that describes the education system's performance in achieving desired educational conditions and outcomes: the indicator is thus linked to the goals of the system and provides a benchmark for measuring progress;
b) information about features known through research to be linked with desired outcomes: such indicators have predictive value because when they change, other changes can be expected to follow;
c) information that describes central features of the system (*e.g.* inputs) in order to understand how the system works;
d) information that is problem-oriented;
e) information that is policy-relevant: indicators should describe educational conditions of particular concern to policy-makers and be amenable to change by policy decisions.

In addition to these substantive criteria, Oakes (1986, p. 2) contends that indicators should have the following technical characteristics:

a) indicators should measure ubiquitous features of schooling that can be found in some form throughout the system, so that information can be compared across diverse settings;
b) indicators should measure enduring features of the system so that trends over time can be analysed;

c) indicators should be readily understood by a broad audience;
d) indicators should be feasible in terms of time, cost and expertise required for data collection;
e) indicators should be generally accepted as valid and reliable statistics.

From these different perspectives, Ruby *et al.* (1989) draw out certain common attributes of education indicators. In seeking to monitor many aspects of human well-being and to assess the impacts of major developments in society, indicators are concerned with measuring end results or ultimate outputs.

Several concepts flow from this output orientation. One is the normative or goal-based character of indicators. Using indicators to measure progress towards a goal often creates controversy, since there is often disagreement about what the goal is and to what extent performance is an improvement or signifies a decline.

The relevance of an indicator depends on a combination of its validity and accuracy. Since educational output tends to be a vague and subjective concept, there will always be some ambiguity about the validity of an indicator focused on such a concept. Also, any education indicator system that omits qualitative descriptions will be a partial system that may distort policy making.

Indicators also need to be representative. An indicator is not simply a measure of itself but acts as a summary variable for some broader concept. Obviously, indicators do not tell everything about education systems. Instead, they provide an ''at a glance'' profile of current conditions. However, there has always been a major conceptual problem with indicators, in that at a broad level they can mask widely divergent situations within sub-groups. Indicators should therefore be capable of disaggregation so as to allow specific examination of sub-populations.

There is little, if any, disagreement with this position in the current literature. As a consequence, much of the unproductive effort spent in search of the ultimate definition of an indicator during the 1970s has been channelled into action. As Nuttall (1992) concludes, the experience of earlier efforts to develop social indicators demonstrated that technical quality is not sufficient to guarantee their use and continuation. Indicator systems must also produce information useful to the policy-makers if they are to survive as publicly supported endeavours. There thus seems little prospect of any radical change in definition in the near future, nor does it seem that one is necessary.

Indicator systems and models

In the literature, much attention is given to explanations of the need for indicator systems and to the closely linked topic of conceptual models for organising the indicators. While a certain consensus has been reached on the definition of education indicators as statistics which reveal something about the health or performance of education, describe its core features and are useful for decision-making, there has been almost universal rejection of the once popular attempts to develop one general or comprehensive index of education equivalent to the GNP – or what Fanchette (1974) describes as some sort of gross happiness product (*cf.* Ferriss, 1969; Drenowski, 1970; Gooler, 1976). While D'Agostino (1974) sees such a comprehensive indicator as providing more stability than single values and a means of reducing the complexity of the data, most writers now concur with Odden (1990) that a single indicator or even a large number of indicators by themselves cannot fully describe the complexities of the schooling process or an educational system.

Nuttall's views (1992) reflect the current consensus that, given the complexity and diversity of education systems, it is obvious that a single indicator conveys limited information. It is necessary to build a system of indicators – *i.e.* a coherent set that provides a valid representation of the condition of the education system – and not just a collection of statistics. Ideally, an indicator system will provide information about how individual components work together to produce an overall outcome. In other words, the policy and interpretive value to be gained from a system of indicators is greater than the sum of its parts.

There appears to be little disagreement with these objectives. The arguments arise over what the elements of the system should be. Shavelson *et al.* (1989) note that:

> "Past experience with social indicators has demonstrated the need for indicator systems to be firmly grounded in a working theory or model of how the social system being measured actually operates." (quoted from de Neufville, 1975)

The model may be very simple and intuitive or it may be complex, but it must represent phenomena of interest and identify its most important components and the relationships among them. Only with such a model can we have a context for interpreting indicators and for exploring the trends reported by the indicator systems.

Anderson (1986), Oakes (1986) and others emphasize the need to construct a model of how the education system works, in order to specify which indicators should be selected. More than any other factor, the model chosen as the basis for indicators will influence what information an indicator system will provide. Scheerens (1992) identifies three trends in the development of these approaches to conceptualising education indicator systems, which mirror to a large extent changes in the emphasis in projected uses for indicators.

The first trend in the development of education indicators is the transition from descriptive statistics (largely input and resource measures) to those attempting to measure performance or outcomes (more generally, a shift towards figures with evaluative importance). This trend reflects the call for policy-makers to know what is happening in their systems and for schools and systems to become more accountable.

The second trend can be characterised by a movement towards more comprehensive indicator systems, with the addition of an output-outcome measure and context measures, and a growing interest in "manipulative – input factors or process characteristics" (see Stern, 1986; Teauber, 1987). This interest follows on from what Kaagan and Coley (1989) observe to be a higher level of intervention in the top management of education by policy-makers in recent years. They want to know not only what is happening but why and what they can do about it. Process indicators generally refer to characteristics of educational systems that can be manipulated. Adding process measures therefore enhances the policy relevance of indicator systems. If process indicators are considered as referring to the procedures or techniques that determine the transition of inputs into outputs, then increased attention will be given to what happens in schools.

The third trend in conceptualising indicators is a concern to measure data at more than one aggregation level (Scheerens, 1992). The conclusion is that the context – input-process/output model – is still the most useful analytic scheme to systematise thinking about indicator systems. Nuttall (1992) points out that in the development of a system of international education indicators, the choice of a suitable model is particularly difficult since the educational and political values of nations vary, as do their institutions and procedures.

Van Herpen (1992) provides an extended review of issues related to conceptual models of education indicators. Selden (1992) also takes up these issues in some depth. The question of which model is most appropriate in which circumstance, and the level of complexity required, is likely to continue to attract attention. As the drive to produce indicators itself leads to improved and more comprehensive data, part of the problem of defining chains of causality and relationships between the elements making up the education process may be better understood.

Potential uses for indicators

The question of how indicators could potentially be used also receives substantial coverage in the literature. De Neufville (1975), Sheldon (1975) and MacRae (1985) saw the overarching purposes of indicators as characterising the nature of a system through its components, their relationships and their changes over time. While these purposes still hold true today, these early writers saw indicators primarily as providing an information base, with social reporting being the key element in their use. Simply providing information, albeit with an evaluative slant is of course still important, but along with growing acceptance in the mid-1980s of the role of indicators as gauging the health of the system came a greater emphasis on accountability of achievement and responsibility for performance.

Most of the literature from the United States at this time was primarily concerned with measuring student performance at the macro level, *i.e.* at system, state or national level. Kaagan and Smith (1985) are typical in proposing that indicators may help educational agencies to further their reform efforts by: *i)* monitoring changes in key variables, such as the quality of teaching and student performance, which would identify impending problems; *ii)* assessing the impact of educational reform efforts; *iii)* encouraging better performance by comparisons with other nations and states; and *iv)* focusing attention on areas or institutions which require improvement.

Other recent publications – Cuttance (1989), Odden (1990) and Ruby *et al.* (1989) – restate the expanded list of uses for indicators given by Oakes (1986) as follows:

a) report the status of schooling;
b) monitor changes over time;
c) explain the causes of various conditions and changes;
d) predict likely changes in the future;
e) profile the strengths and weaknesses of the system;
f) inform policy-makers of the most effective ways to improve the system;
g) inform decision-making and management;
h) define educational objectives.

Cuttance (1989) discusses these possible uses at some length. Whilst all appear to be theoretically possible, Oakes (1986) considers that only some may be achievable and that others are ultimately unrealistic. This theme is also taken up by others and has helped to focus attention on the practical application of indicators. Shavelson *et al.* (1989), for example, consider that they cannot be used to set goals and priorities, since even though they can inform about these objectives, they are only one factor in the decision-making process; nor can they be used to evaluate programmes – they do not provide the level of rigour or detail necessary – or to develop a balance sheet, since social indicators lack the common referent available to economic indicators. What they can do is to describe and state problems more clearly, signal new problems more quickly, and obtain clues about promising new endeavours.

Two other shifts in focus from the earlier literature can be detected. Not only is there a suggestion that indicators can do more than simply provide information in a form which allows comparison (for readers to make up their own minds about what indicators mean and what should be done – a theme particularly strong in the United Kingdom), but that they can and should play a more active part in the process.

This view is reflected in three ways. The first, which is an extension of the previous accountability orientation, is the overt linking of performance on certain indicators to rewards and sanctions (Honig, 1985; Pipho, 1988; Kaagan and Coley, 1989). The second, which might be characterised as an orientation towards improvement, is the more subtle use in focusing attention on what is important and worth doing – see Porter (1988) and David (1988) for discussion of political influences on indicator selection and of possible threats to local control; and the critics of minimum competency testing, *e.g.* Haney and Madaus (1978); Wise (1979); Madaus (1981); Stedman and Kaestle (1985); Gray *et al.* (1986); Gipps (1988); and Goldstein and Cuttance (1988). Thirdly, indicator systems in education both influence and are themselves influenced by the social system in which they exist.

These potential uses have more often been envisaged for system-level use or for what Oakes (1986) describes as "top-down" development. In the past few years there has been a growing body of concern with the "bottom-up" use of indicators at the school level, with a focus on school improvement through self-evaluation, including better planning and financial management. Anderson (1986) argues that there are both statistical and practical reasons why the focus of attention should be as close as possible to where the service is delivered, *i.e.* at the school or even classroom level.

Scheerens (1992) discusses how school process indicators might be derived from the findings of the literature on effective schools (and how these might then be of value in contributing to national-level indicators). Another approach which shows some promise sees indicators forming a check-list against which schools can evaluate themselves (Wakefield, 1988), or be examined by outside agencies such as an inspectorate (Cuttance, 1989). A further refinement places greater emphasis on the process of indicator development than on the indicators themselves, which forces schools and their communities to clarify their goals and reflect on their outcomes (Marshall, 1988; Ashenden, 1987b; Ruby and Wyatt, 1988). Caldwell and Spinks (1986) also make extensive use of indicators in the programme budgeting cycle – when attention has to be paid to how inputs contribute to processes and outputs – as a means of creating self-managing schools.

One final use for indicators concerns their application to staff appraisal schemes. Those authors who do address this issue place it in the context of revivals of "payment by results" or, more usually, "merit pay" in the United States (Murnane and Pauly, 1988). McCormack and Stevens (1990) describe a pilot project to develop indicators for principals of Catholic schools.

In summary, education indicators may be used in four main ways: to meet accountability requirements; to defend or legitimate what has been done; to help in decision-making; and to focus effort, such as setting targets for achievement. The particular emphasis given to these uses is still emerging and varies from country to country. It seems unlikely that anything other than refinements to them will be made in the near future. But as these ideas become more widely implemented, the causes of success or failure of various schemes will be examined more fully and more analytically, leading to a clearer understanding of what is both heuristically and conceptually possible. Looking to the future, it appears that meeting accountability requirements will continue to be the dominant theme, and therefore this may need to be balanced by attention to indicators for developmental and planning purposes.

Indicators in action

The literature describing the application of indicators falls into three main categories: regular publications providing indicator data, those describing work in progress, and those analysing these efforts.

Alkin (1988) describes indicator systems in New Guinea, Venezuela, the Philippines, Indonesia, Israel, Canada and Scotland; but it has usually been developments in the United States that have received most attention. Griffith (1990) describes the "flagship" publications of the U.S. National Center for Education Statistics – *The Condition of Education*, *Youth Indicators*, and the third indicator publication of the U.S. Department of Education, the Wall Chart. Many writers have commented on the substance of the Wall Chart, particularly in relation to the reporting of Scholastic Aptitude Test (SAT) scores (Powell and Steelman, 1984).

California has been a leader in the United States in education indicator initiatives. Odden (1990) discusses three of the California indicator reports – *Conditions of California Education*, published by Policy Analysis for California Education (PACE); the Quality Indicators Reports of the State Education Department; and the compulsory School Accountability Report Cards kept by Californian schools. While policy imperatives have been the major factor behind many of these indicator projects, over the years California has created a data infrastructure that allows it to produce indicator systems. It has a comprehensive school-by-school data system that can provide detailed information on students, teachers, school context and curriculum, instruction and student performance. Without this kind of data, a comprehensive indicator system cannot be produced.

Inman *et al.* (1990) provide the most recent and comprehensive review of state education indicator systems which recognises that the field is a rapidly expanding one (see also Pipho, 1988). For each state, a profile has been drawn up which identifies the legislative base or reform initiative responsible for the indicator scheme, major publications reporting the indicators, and any development likely to occur in the future. For example, each school district in Illinois must submit a school report card assessing the performance of its schools and students. The card, an index of school performance measured against statewide and local standards, and providing information to enable annual comparisons and set future targets, includes the following indicators:

- district and statewide student performance;
- percentage of students in top and bottom quartiles of national-norm achievement tests;
- composite and subtest means for college-bound students;
- student attendance rates;
- number of chronic truants;
- percentage of students not promoted to next grade;
- graduation rates;
- student mobility;
- class size;
- percentage of enrolments in high school mathematics, science, English and social science;
- amount of time devoted daily to mathematics, science, English and social science;
- pupil/teacher ratios;
- operating expenditure per pupil;
- per capita tuition charge;
- district expenditure;
- administrators' average salary;
- teachers' average salary.

From their study of 50 states, Kaagan and Coley (1989) conclude that indicator results are being used prematurely to hold local schools accountable, that there is insufficient investment to ensure high quality of the measuring that becomes part of the state indicator system; and that there is a reluctance to understand the indicators, *i.e.* to postulate a relation between inputs and outcomes for the purpose of recommending policy action.

Odden (1990) underlines the need to be clear about the purpose of an education indicator system and not to use it inappropriately. Both Kaagan and Coley (1989) and McDonnell (1989) argue that using indicators for accountability is still premature because few systems yet provide enough information and there is still too much to learn about the linkages of cause and effect between indicators. Current systems also provide data too far removed from the classroom level to improve local conditions. The authors conclude that indicator systems are therefore best used to offer policy-makers a broad overview of the education system.

Selden (1990) finds that almost all the state indicator systems reviewed now go beyond mere reporting of student achievement testing programmes. They reflect in varying degrees recent professional developments as well as individual histories and circumstances of the states, and the balance between state and local control of policy-making. Selden notes that some indicator systems seem to instigate public interest – through publicity surrounding test-score results, for example – in pressing schools and school districts to improve. Some seem designed to reward good schools and districts and to identify low-performing ones. Other systems provide state-level measures of quality as evidence of effectiveness of state-level reform, or provide insights into the workings of educational programmes.

Selden goes on to identify four trends in the development of state indicator systems. First, he believes that the trend towards their centralisation appears to be virtually complete. Even those states lacking indicator programmes report on comparative achievement data for local schools and districts, or are beginning to establish them. Second, in

BEST COPY AVAILABLE

fewer than ten states are reports not part of an integrated indicator system. Third, the systems do not reflect convergence on common approaches or models. Finally, the earlier trend towards more extensive indicator systems with direct policy consequences appears to be levelling off.

In contrast to these "top-down" accountability systems in the United States are the "bottom-up" developments in Australia, New Zealand, the United Kingdom, and elsewhere. Wyatt (1988a and 1990) provides an overview of these developments within government school systems – see Ashenden (1987*a*) and Marshall (1988) for a discussion of state-level initiatives. The form of accountability appears to be influenced by the local environment, with the Australian and New Zealand schemes, for example, attempting to link accountability with improvement in the quality of schooling by providing information that will assist school-level administrators.

Hocking and Langford (1990), Irving (1990) and Osborne (1990) all provide useful information on the practical arrangements and the issues that have shaped them in three different education systems in Australia and New Zealand. Concern for equity has also been a major focus: Shrubb (1990) describes a project to develop indicators for gender equity; Millan (1990) outlines a proposal to monitor the outcomes of special-purpose programmes for disadvantaged schools and for aboriginal students; and Keefe (1990) provides a provocative analysis of ideological resistance to evaluation of outcomes in aboriginal education. He stresses the need for strategies such as institutional profiles and indicators to concentrate efforts on the achievement of equity goals. Stevens (1990) gives details of the approach to developing indicators that combines school-level and system-level evaluation of a programme to increase participation in schooling (see Lawton *et al.*, 1988, for a review of the potential role for indicators to monitor student retention and transition in Ontario high schools).

All this work is important. It shows that indicators and related techniques can go past the point of theory to application. There is much to be learned from these accounts, since they offer insights into how to introduce and institutionalise indicators and other evaluation techniques effectively and identify pitfalls to be avoided. In addition, they can provide a basis for assessing the efficacy of more general applications.

Equally important is the contribution these experiences can make to the theoretical work on indicators. At a technical level, it is apparent that practitioners, and others, are moving away from definitions when constructing individual indicators, causing the emerging methodology to be confused with education statistics, management by objectives, and general evaluation strategies. This tendency could also reduce the benefits that flow from investment in indicators if they do not serve their policy and monitoring ends.

Developing improved indicators

Almost all of the writers who have contributed to the conceptual development of indicators have mentioned various problems and limitations. Their experiences are not confined to education but are also found in other areas including the electricity industry (Curran, 1988), the airline industry (David, 1988), other nationalised industries (Woodward, 1986) and other public service areas (Bourne, 1984). These problems can be summarised as follows (Ruby and Wyatt, 1988):

a) indicators provide limited information;

b) there can be problems with simple models;
c) there are problems with the collection and analysis of indicator data;
d) indicators affect subsequent performance;
e) indicators create political pressures.

To these Oakes (1986) adds the following:

f) the level of information collected;
g) the challenges of making comparisons;
h) the costs and benefits of extensive indicator systems;
i) the political pressures that the existence of an indicator system will bring;
j) the vital question of who makes the design decisions for any indicator systems.

Various commentators examine these issues in greater depth. David (1988), Porter (1988) and Garbutcheon-Singh (1989) discuss some of the political implications of indicator systems, in terms of possible effects on local control over teaching, standardisation of the curriculum, and discouragement of innovation. They point to the risk of staff time and money being spent on collecting data which will then not be used, and of indicators becoming ends in themselves (*cf.* Wood and Power, 1984).

Consideration of conceptual problems generally receives less coverage, Eide (1987) and Shavelson *et al.* (1989) notwithstanding, although information useful for both operationalising indicator development and shaping conceptual frameworks has been generated from comparisons with earlier work on social indicators (Rockwell, 1990) and economic indicators (Murnane, 1987; Horn and Winter, 1990). Understanding the nature of these common problems helps in appreciating the progress that has been made in improving education indicators. Murnane concludes that in both cases it is very difficult to develop operational measures of performance that are closely related to their desired outcome. There are always trade-offs to be made in determining the level of disaggregation for reporting performance measures. It is, moreover, difficult to determine sub-group breakdowns which are the most likely to be helpful in explaining performance trends.

The comparison between economics and education indicators illustrates that many of the problems in developing social indicators are generic and not unique to education. Recognition of this general principle may encourage analysts working to improve education indicators to learn from the experience of indicator development in other sectors.

Murnane and Pauly (1988) draw three lessons from this experience. Firstly, it is important to develop multiple measures. No single set of test scores offers a workable basis for assessing the performance of the education system, just as no single measure – such as the unemployment rate – provides a reliable measure of the "health" of the economy. Secondly, indicators make us aware of new and puzzling questions. They do not provide all the answers, nor can they be expected to. Thirdly, it is important to educate the users of indicators. In the economic field, a rich set of indicators has been available for over 40 years, and users have become well versed in piecing together a picture of economic performance by examining trends in more than one indicator. The development of a comparable set of education indicators is still evolving, and most users are still learning how to use them and how to interpret the patterns appearing in the data.

Horn and Winter (1990) point out that the most successful economic indicators in use today have a well-defined purpose in the policy formulation process and are linked to specific policies, programmes or legislation. When, for example, payment of millions of dollars for welfare programmes is tied to a particular indicator (such as the CPI), there is substantial pressure to justify its validity. Furthermore, modern economic indicators have developed interactively over a long time and it is unlikely that a new indicator could come into practice in less than a decade.

Economic indicators have gained the support of the research community in part because they are based on widely accepted methodological procedures and sophisticated survey techniques, and their validity and reliability has been established in different ways. The system is therefore not subject to the influence of individual agencies or data-reporting entities. In addition, substantial financial resources have been and continue to be invested in developing economic indicators. Considerable political support to ensure the resources necessary for the development and maintenance of a new indicator is usually required.

Avoiding overstatement of both short-term and long-term benefits of indicators is also important. Although they can have a large part to play in both the planning process and accountability schemes, indicators are by no means the sole answer to an organisation's problems. They can at best only point to the existence of problems or successes and cannot suggest solutions. The technical limitations to the construction and interpretation of indicators also need to be recognised.

The literature on technical problems can be grouped into two main sub-categories: those addressing the issue of how to make fair comparisons, and those concerned with the adequacy of available measures. The literature on each of these areas is quite substantial, but while all have some relevance to the question of how to develop better indicators and indicator systems, not all directly address the concept of indicators themselves.

Hanushek (1970), Benveniste (1984), Garms (1984), Guthrie and Kirst (1984), Goldstein and Cuttance (1988), Burstein *et al.* (1989) and Nuttall (1992) report on methods of judging the relative effectiveness of schools. There is a move away from earlier "standards" models to those which attempt to measure the "value added" by schooling. The discussion focuses on how changes over time can be measured, whether and how differences in student performance should be adjusted to account for influences not due to the effects of schooling, and how to deal with the multilevel nature of educational processes.

The sophisticated statistical techniques of these methods lose much of their power if the basic data they require are inadequate. A long-standing criticism (Alexander and James, 1987) of many of the early (and existing) national and state accountability schemes is that indicators of student outcomes have been limited in the areas they cover and the skills they measure, and that they are lacking in diagnostic value. Fairly good paper and pencil tests of the most commonly taught basic knowledge and skills exist, but adequate measures of children's ability to think critically, to apply knowledge and to solve problems are lacking (Oakes, 1986). These shortcomings have, in part, led to attempts to make greater use of teacher-assessed tasks in the national assessment scheme in the United Kingdom (Black, 1988).

Not only student assessment lacks adequate measures. Oakes (1986) notes that while most agree that the quality of the teaching force should be a key indicator, teacher quality can at present only be perceived through indirect measures such as qualifications and years of experience – factors which correlate only weakly with student outcomes. It remains to be seen whether this reliance on "proxy" measures can be overcome or will remain an intrinsic weakness in indicator models.

A further issue of concern to those attempting to develop state, national and internationally comparable indicators is the lack of consistency in the definitions, and thus common understanding of even the most basic concepts such as "What is a school?" and "Who is a drop-out?". This has been a recurrent theme in each of the OECD sponsored conferences on indicators (Suter and Sherman, 1990; Ruby, 1992), and one which shows little sign of resolution without a great deal of compromise (Burstein *et al.*, 1989).

Whereas analyses of the weaknesses of indicator systems seldom directly challenge their conceptual basis, a deeper problem, not often addressed, concerns the proper role of indicators in the organisation of schooling. The managerial model dominant today is a relative newcomer to public administration and has yet to prove more successful than the previous bureaucratic and professional models. Fasano, in Chapter 3 of this volume, explores this issue further.

Conclusion

This chapter has attempted to trace some major trends in the development of performance indicators in education. It has shown that the concept is both old and new, and that, in its most recent phase it has proved to be both dynamic and productive. Yet it should legitimately be asked what, if anything, is new, and what has changed since the concept re-emerged. Even though the literature is prolific, certain key figures and developments have played a more important part than others in furthering our understanding of indicators, and thus recur frequently in this review.

Further work on definitions, and continuing analysis of attempts to apply indicators, will clearly be important areas of investigation in the future. In addition, an examination of work in progress should address the pressing question of whether and how the outcomes of the emerging work at the school level can be aggregated and selectively assembled into an indicator system capable of representing system-level interests. This type of analysis would underpin the larger-scale developmental work on models of education required for a comprehensive indicators system.

Related to this work is the need for a critical examination of the models of accountability implicit in definitions of indicators and in the shape and structure of indicator systems, in order to define more rigorously the relationships between accountability, quality and indicators. Greater attention to the data collection and selection issues that need to be resolved before indicator systems can be fully implemented is also necessary. Work on standardization and the compatibility of elements within education data systems needs to be undertaken, and there are major data management questions that need to be resolved.

Without continued research and analysis, the considerable benefits promised by indicators will not be realised. In the meantime, indicators seem set to remain a "hot" topic in the education literature.

Bibliography

ALBERTA MINISTRY OF EDUCATION (1990), *Educational Quality Indicators: Annotated Bibliography* (2nd edition), Policy and Planning Secretariat/Library Services, Ministry of Education, Alberta, Canada, January.

ALEXANDER, L. and JAMES, H.T. (1987), *The Nation's Report Card: Improving the Assessment of Student Achievement*, U.S. National Academy of Education, Stanford, CA.

ALKIN, M.C. (1988), "National Quality Indicators: A World View", *Studies in Educationai Evaluation*, Vol. 14 (1), pp. 11-24.

ANDERSON, P.S. (1986), *Beyond the Wall Chart: Issues for States*, Office of Educational Research and Improvement, U.S. Department of Education, Washington, D.C.

ASHENDEN, D. (1987*a*), "Indicators, Information and Education: A Discussion of Possibilities and Problems". Prepared for the Curriculum Development Council, Canberra, September.

ASHENDEN, D. (1987*b*), "Information and Inequality: A Systematic Approach to Increasing Educational Equality", Working paper, Social Justice Standing Committee, Victoria, Australia, September.

ASHENDEN, D. (1987*c*), "An Odd Couple/Social Justice and Performance Indicators", Working paper, November, Social Justice Standing Committee, Victoria, Australia.

AUSTRALIAN BUREAU OF STATISTICS (1976), *Social Indicators*, Catalogue No. 4101.0, Australian Bureau of Statistics, Canberra.

BAUER, R. (ed.) (1966), *Social Indicators*, MIT Press, Cambridge, MA.

BENVENISTE, G. (1984), "The Design of School Accountability Systems", in Guthrie, J.W. and Kirst, M.W. (eds.), *Data-Based Accountability in Education*, Policy Analysis for California Education (PACE), Stanford, CA.

BLACK, J. (1988), *National Curriculum: A Report of the Task Group on Assessment and Testing*, Department of Education and Science, Welsh Office, London.

BOTTANI, N. and DELFAU, I. (eds.) (1990), "Indicators of the Quality of Education Systems: An International Perspective", *International Journal of Educational Research*, Vol. 14 (4), pp. 321-408.

BOURNE, J.B. (1984), "Performance Appraisal in the Public Sector", *Public Administration*, Vol. 60 (3).

BROAD, J. (1983), *Education Indicators: A Review of the Literature*, Victoria Education Department, Canberra.

BURSTEIN, L., OAKES, J. and GUITON, G. (1989), "Stages in the Construction of Indicator Systems for Assessing Educational Quality". Presented at the Symposium on Developing State Indicator Systems: From Theory to Practice. American Educational Research Association Annual Meeting, San Francisco, CA, March.

CALDWELL, B.J. and SPINKS, J.M. (1986), *Policy-Making and Planning for School Effectiveness*, Tasmania Education Department, Hobart.

112

CAMPBELL, A. and CONVERSE, P.E. (eds.) (1972), *The Human Meaning of Social Change*, Russell Sage Foundation, New York.

CARLEY, M. (1981), *Social Measurements and Social Indicators*, George Allen and Unwin, London.

CURRAN, S. (1988), "Setting Performance Indicators: The Right Deed for the Wrong Reason", in Wyatt, T.J. and Ruby, A. (eds.), *Indicators in Education*, Australian Conference of Directors-General of Education, Sydney, pp. 55-66.

CUTTANCE, P.F. (1989), *Performance Indicators for Schooling: A Report to the Scottish Department of Education*, Department of Education, Edinburgh.

D'AGOSTINO, R.B. (1974), "Social Indicators: A Statistician's View", *Social Indicators Research*, Vol. 1, pp. 459-544.

DAVID, J.L. (1988), "The Use of Indicators by School Districts: Aid or Threat to Improvement?", *Phi Delta Kappan*, Vol. 69 (7), pp. 449-450.

DE NEUFVILLE, J.I. (1975), *Social Indicators and Public Policy: Interactive Processes of Design and Application*, Elsevier Scientific Publishing Company, New York.

DRENOWSKI, J. (1970), *Studies in the Measurement of Levels of Living and Welfare*, Report No. 70, UNRISD, Geneva.

EIDE, K. (1987), "The Need for Statistical Indicators in Education". Presented at the OECD International Conference on the Evaluation of Education Systems, Washington, D.C., November.

FANCHETTE, S. (1974), "Social Indicators: Problems of Methodology and Selection", in *Social Indicators: Problems of Definition and of Selection*, UNESCO, Paris.

FERRISS, A.L. (1969), *Indicators of Trends in American Education*, Russell Sage Foundation, New York.

FETLER, M. (1986), "Accountability in California Schools", *Educational Evaluation and Policy Analysis*, Vol. 8 (1), pp. 31-44.

FRACKMAN, E. (1987), "Lessons to be Learnt from a Decade of Discussions on Performance Indicators", *International Journal of Institutional Management in Higher Education*, Vol. 11 (2), pp. 149-162.

FUHRMAN, S. (ed.) (1988), "An Overview", *Phi Delta Kappan*, Vol. 69 (7), pp. 486-487.

GARBUTCHEON-SINGH, M. (1989), "Indicters or Vindicators of Inequity", *Education Links*, Vol. 36, p. 18.

GARMS, W.I. (1984), "Merit Schools for Florida: A Concept Paper", in Guthrie, J.W. and Kirst, M.W. (eds.), *Data-Based Accountability in Education*, Policy Analysis for California Education (PACE), Stanford, CA.

GIPPS, C. (1988), "Standards in Education and the Role of Testing", *British Journal of Educational Studies*, Vol. 36 (1), pp. 21-36.

GOLDSTEIN, H. and CUTTANCE, P. (1988), "A Note on National Assessment and School Comparisons", *Journal of Education Policy*, Vol. 3 (2), pp. 197-202.

GOOLER, D.D. (1976), "The Development and Use of Educational Indicators", *Educational Indicators: Monitoring the State of Education*, Educational Testing Service, Princeton, NJ.

GRAY, J., JESSON, D. and JONES, B. (1986), "The Search for a Fairer Way of Comparing Schools: Examination Results", *Research Papers in Education*, 1.

GRIFFITH, J. (1990), "Indicators and Accountability in the United States", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 1-14.

GUTHRIE, J.W. and KIRST, M.W. (1984), *Data-Based Accountability in Education*, Policy Analysis for Californian Education (PACE), Stanford, CA.

HANEY, W. and MADAUS, C. (1978), "Making Sense of the Competency Testing Movement", *Harvard Educational Review*, Vol. 48 (4), pp. 462-484.

HANUSHEK, E.A. (1970), *Developing Local Education Indicators – The Priorities*, Rand Corporation, Santa Monica, CA.

HOCKING, H. and LANGFORD, J. (1990), "Accountability and Quality: A School-based Perspective", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 191-204.

HONIG, W. (1985), *Performance Report for California Schools*, California State Department of Education, Los Angeles, CA.

HORN, R. and WINTER, C. (1990), "Common Factors in the Development of Economic Indicators: Lessons Learned". Prepared for the National Center for Education Statistics (NCES), Indicators Panel, Washington, D.C., October.

INMAN, D., PRIBESH, S.L. and SALGANIK, L.H. (1990), "Summary Profiles, State Education Indicator Systems". Prepared for the National Center for Education Statistics (NCES) Indicators Panel, Washington, D.C., October.

IRVING, J. (1990), "Plans for Quality Review and Accountability Mechanisms in New Zealand", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 95-112.

JAEGER, R.M. (1978), "About Educational Indicators: Statistics on the Conditions and Trends in Education", *Review of Research in Education*, pp. 276-315.

JOHNSTONE, J.N. (1981), *Indicators of Education Systems*, Kogan Page, London; UNESCO, Paris.

KAAGAN, S. and COLEY, R. (1989), *State Education Indicators: Measured Strides, Missing Steps*, Rand Corporation, Santa Monica, CA.

KAAGAN, S. and SMITH, M.S. (1985), "Indicators of Educational Quality", *Educational Leadership*, October, pp. 21-24.

KEEFE, K. (1990), "Wet Left and Dry Left: Performance Indicators for Aboriginal Students in Higher Education", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney.

LAND, K.C. (1975), "Theories, Models and Indicators of Social Change", *International Social Science Journal*, Vol. 27 (1), pp. 7-37.

LAWTON, S., LEITHWOOD, K.A., BATCHER, E., DONALDSON, E.C., and STEWART, R. (1988), *Student Retention and Transition in Ontario High Schools: Policies, Practices and Prospects*, Ontario Ministry of Education, Ottawa.

MacRAE, D., Jr. (1985), *Policy Indicators: Links Between Social Science and Public Debate*, University of North Carolina Press, Chapel Hill, NC.

MADAUS, G.M. (1981), "Minimum Competency Testing. A Critical Overview". Presented at the 46th Annual Conference of the Education Records Bureau, New York, October.

MARSHALL, G. (1988), "The Application of Performance Indicators at the School Level to a Programme Planning and Evaluation Cycle", in Wyatt, T.J. and Ruby, A. (eds.), *Indicators in Education*, Australian Conference of Directors-General of Education, Sydney, pp. 75-88.

McCOLLUM, H. and TURNBULL, B.J. (1990), "Education Indicators". Prepared for the National Center for Education Statistics (NCES) Indicators Panel, Washington, D.C., October, pp. 179-190.

McCORMACK, T. and STEVENS, P. (1990), "The Development and Evaluation Programme for Principals in Tasmania Catholic Schools: Some Performance Indicators", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 179-190.

McDONNELL, L.M. (1989), "Coursework Policy in Five States and its Implications for Indicator Development", Working paper, Rand Corporation, Santa Monica, CA.

MILLAN, M. (1990), "Indicators and Equity in Education: Aboriginal Education and the Disadvantaged Schools Programme". Addendum to Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney.

MURNANE, R.J. (1987), "Improving Education Indicators and Economic Indicators: The Same Problems?", *Educational Evaluation and Policy Analysis*, Vol. 9 (2), pp. 101-116.

MURNANE, R.J. and PAULY, E.W. (1988), "Lessons from Comparing Educational and Economic Indicators", *Phi Delta Kappan*, March, Vol. 69 (7), pp. 509-513.

NATIONAL CENTER FOR EDUCATION STATISTICS (1985), *Indicators of Education, Status and Trends*, U.S. Department of Education, Washington, D.C.

NATIONAL CENTER FOR EDUCATION STATISTICS (1987), *The Condition of Education: A Statistical Report*, U.S. Department of Education, Washington, D.C.

NATIONAL CENTER FOR EDUCATION STATISTICS (1989), *Youth Indicators*, U.S. Department of Education, Washington, D.C.

NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION (1983), *A Nation at Risk: The Imperative for Educational Reform*, U.S. Government Printing Office, Washington, D.C.

NUTTALL, D. (1992), "The Functions and Limitations of International Education Indicators", *The OECD International Education Indicators. A Framework for Analysis*, OECD/CERI, Paris, pp. 13-23.

OAKES, J. (1986), *Education Indicators: A Guide for Policymakers*, Rand Corporation, Santa Monica, CA.

ODDEN, A. (1990), "Making Sense of Education Indicators: The Missing Ingredient", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 33-50.

OECD (1973), "A Framework for Educational Indicators to Guide Government Decisions", OECD, Paris (document for general distribution).

OECD (1989), *R&D Statistics and Output Measurement in the Higher Education Sector, Supplement to the Frascati Manual*, OECD, Paris.

OSBORNE, D. (1990), "The NSW School Development and Evaluation Model", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 147-154.

PIPHO, C. (1988), "The Uses of Test Scores", *Phi Delta Kappan*, Vol. 70 (4), pp. 278-279.

PORTER, A. (1988), "Indicators: Objective Data or Political Tool?", *Phi Delta Kappan*, Vol. 69 (7), pp. 503-508.

POSTLETHWAITE, T.N. and WILEY, D.E. (eds.) (1992), *The IEA Study of Science II: Science Achievement in 23 Countries*, Pergamon Press, Oxford.

POWELL, B. and STEELMAN, L.C. (1984), "Variations in State SAT Performance: Meaningful or Misleading", *Harvard Educational Review*, Vol. 54 (4), pp. 389-412.

ROCKWELL, R.C. (1990), "Lessons from the History of the Social Indicators Movement". Prepared for the National Center for Education Statistics (NCES) Indicators Panel, Washington, D.C., October.

RUBY, A. (1989), "Why Invest in Performance Indicator Development?", in Ruby, A. *et al.*, *Reporting Educational Progress: Performance Indicators in Education. Monograph Series*, Australian Conference of Directors-General of Education, Sydney.

RUBY, A. (1992) "Do Common Values Produce Common Indicators?", in *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, pp. 77-82.

RUBY, A. and WYATT, T.J. (1989), "Using Indicators for School Improvement", in Ruby, A. *et al.*, *Reporting Educational Progress: Performance Indicators in Education*, Australian Conference of Directors-General of Education, Sydney.

RUBY, A., WYATT, T.J., NORTON, S., DAVIES, B. and SHRUBB, S. (1989), *Reporting Educational Progress: Performance Indicators in Education*, Australian Conference of Directors-General of Education, Sydney.

SCHEERENS, J. (1992), "Process Indicators of School Functioning", in *The OECD International Education Indicators. A Framework for Analysis*, OECD/CERI, Paris, pp. 53-76.

SELDEN, R.W. (1990), "State Indicator Systems in Education". Prepared for the National Center for Education Statistics (NCES) Indicators Panel, October, Washington D.C.

SELDEN, R.W. (1992), "Standardizing Data in Decentralised Educational Data", in *The OECD International Education Indicators. A Framework for Analysis*, OECD/CERI, Paris, pp. 107-113.

SHAVELSON, R., McDONNELL, L.M., OAKES, J. and CAREY, N. (1989), *Indicator Systems for Monitoring Mathematics and Science Education: A Sourcebook*, Rand Corporation, Santa Monica, CA.

SHELDON, E.B. (1975), "The Social Indicators Movement", *Educational Indicators: Monitoring the State of Education*, Educational Testing Service, Princeton, NJ.

SHELDON, E.B. and MOORE, W.E. (eds.) (1968), *Indicators of Social Change: Concepts and Measurements*, Russell Sage Foundation, New York.

SHRUBB, S. (1990), "Education Indicators for Gender Equity", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 155-166.

SMITH, M.S. (1984), "A Framework for the Development of National Education Indicators". Prepared for the Council of Chief State School Officers, U.S. Department of Education, Washington, D.C.

SMITH, M.S. (1988), "Educational Indicators", *Phi Delta Kappan*, Vol. 69 (7), pp. 487-491.

STEDMAN, L.C. and KAESTLE, C.F. (1985), "The Test Score Decline is Over: Now What?". *Phi Delta Kappan*, Vol. 67 (3), pp. 204-210.

STERN, J.D. (1986), *The Education Indicators Project in the United States*, Office of Educational Research and Improvement, U.S. Department of Education, Washington, D.C.

STERN, J.D. and HALL, P.R. (1988), *Education Development at the Federal Level*, Office of Educational Research and Improvement, U.S. Department of Education, Washington, D.C.

STEVENS, A. (1990), "Education Indicators and 'Staying On' – A School Effectiveness Programme", in Wyatt, T.J. (ed.), *Indicators and Quality in Education*, Australian Conference of Directors-General of Education, Sydney, pp. 41-46.

SUTER, L.E. and SHERMAN, J.D. (1990), "International Indicators: Current Status and Future Prospects". Prepared for the National Center for Education Statistics (NCES) Indicators Panel, Washington, D.C., October.

TEAUBER, R.C. (ed.) (1987), "Education Data System Redesign". Special issue of *International Journal of Educational Research*, Vol. 11 (4), pp. 391-511.

UNESCO (1980), *Evaluating Research and Social Change*, UNESCO, Paris.

UNITED KINGDOM CENTRAL STATISTICAL OFFICE (1970 and 1971), *Social Trends*, Bulletin, HMSO, London.

UNITED STATES DEPARTMENT OF HEALTH, EDUCATION AND WELFARE (1969), *Toward a Social Report*, U.S. Government Printing Office, Washington, D.C.

van HERPEN, M. (1992), "Conceptual Models in Use for Education Indicators", in *The OECD International Education Indicators. A Framework for Analysis*, OECD/CERI, Paris, pp. 25-51.

WAKEFIELD, B. (1988), "Performance Indicators in Secondary Schools: Some Practical Considerations". Presented at the National Foundation for Educational Research (NFER) briefing, Conference on Quality in Schools, London, June.

WISE, A.E. (1979), *Legislated Learning: The Bureaucratization of the American Classroom*, University of California Press, Berkeley, CA.

WOOD, R. and POWER, C. (1984), "Have National Assessments made us any Wiser about Standards?", *Comparative Education*, Vol. 20 (3), pp. 307-321.

WOODWARD, S.N. (1986), "Performance Indicators and Management Performance in Nationalized Industries", *Public Administration*, 64, Autumn.

WYATT, T.J. (1988*a*), "Recent Developments in Performance Indicator Implementation in Australian School Systems", Presented at the National Evaluation Conference, Melbourne, July.

WYATT, T.J. (1988*b*), "What can be Learned from the Social Indicators Movement?", in Wyatt, T.J., Small, J. and Norton, S. (eds.), *Education Indicators: Bibliograghy. Reporting on Educational Progress: Performance Indicators in Education*. Australian Conference of Directors-General of Education, Sydney.

WYATT, T.J. (1990), "Indicators for School Improvement: Lessons from the First 12 Months", in Wyatt, T.J. and Ruby, A. (eds.), *Education Indicators for Quality, Accountability and Better Practice*, Australian Conference of Directors-General of Education, Sydney, pp. 139-146.

WYATT, T.J. and RUBY, A. (eds.) (1988), *Indicators in Education*, Australian Conference of Directors-General of Education, Sydney.

WYATT, T.J., SMALL, J. and NORTON, S. (eds.) (1988), *Education Indicators: Bibliography. Reporting on Educational Progress: Performance Indicators in Education*. Australian Conference of Directors-General of Education, Sydney.

Chapter 6

Developing and Implementing Education Indicators

by

Rolf K. Blank
Council of Chief State School Officers, United States

The OECD has recently engaged in efforts to construct a system of international education indicators, which are likely to lead to plans for collecting and reporting data on the selected indicators (Bottani and Walberg, 1992). Lessons learned from other efforts to develop education indicators, such as the experience of state-by-state education indicators in the United States, can be useful in considering key issues for implementing indicators and in identifying important steps for building a co-operative system for collecting and reporting data.

After participating countries have selected and agreed on a set of indicators, they will need a detailed plan specifying how comparative data will be collected, analysed, and reported. When countries have co-operated to select and define common education indicators, they will have to co-operate further to decide on a methodology for obtaining the data that will be used to measure and report on country-by-country comparative indicators. In 1985, the Council of Chief State School Officers began a process of developing state-by-state education indicators in the United States, which has resulted in an annual report of state education indicators (CCSSO, 1989). The development of state indicators has some lessons to teach for international education indicators (Selden, 1990), and the present chapter provides information concerning issues, strategies, and specific steps for developing valid, comparative indicators based on the CCSSO development process (Blank, 1989; Blank and Dalkilic, 1990).

The chapter has three sections: first, a discussion of issues for developing education indicators involving different governments; second, a description of strategies for developing a set of education indicators; and third, an overview of the main steps for constructing indicators through a co-operative data system.

Issues in developing education indicators

Political, cultural, and organisational differences in the education systems of countries must be considered when selecting education indicators. These differences are likely to shape plans for their implementation. Several key elements will help provide reliable, valid comparisons among countries:

- comparability and data specification;
- commitment, cost, and follow-through;
- analysis and interpretation of data;
- standardized procedures and reporting standards.

Comparability and data specification

Moving from consensus among country representatives on a desired set of indicators to specification of a measure, or data element, for each indicator is a basic problem. Once representatives from different governments have agreed to a set of indicators, they may find that the data to measure and compare education on the desired indicators may or may not be available in their countries. New data collection may be needed, some countries may have to adapt their data systems, or data may be available but not comparable. An agreed set of indicators is likely to be only partly based on currently available data.

In developing state-by-state indicators in the United States, the CCSSO found that valid, comparable data were not available for several of the indicators initially selected (CCSSO, 1985). For the student achievement indicator, for example, many states had achievement tests in the same subjects and grade levels, but the variety of items used in the tests prevented valid comparisons. However, the Chief State School Officers' 1985 decision to work towards a state-by-state indicator of student achievement was a major factor in a decision by the National Assessment of Educational Progress (NAEP) to include representative state samples starting in 1990. Although the development process was lengthy, the Chiefs' agreement on the indicator of student achievement was important.

There are three broad methodological options for collecting and reporting data for international education indicators:

- each participating country uses the same data collection instrument;
- each country has the option of using its own data collection instrument to implement common standards for categorising and reporting data;
- each country reports descriptive information within categories defined for the desired indicator.

The options are displayed in Table 6.1. The first two provide means of gathering and reporting quantified data that can be statistically compared among countries.

The first option is used to assess and compare student achievement in different education systems and is employed, for example, by the International Association for the Evaluation of Educational Achievement (IEA) and National Assessment of Educational Progress (NAEP). Other indicators also require use of a single instrument and set of items, particularly those based on respondents' knowledge, opinions, interpretation, or evaluation. A common data collection instrument (or common items or data elements) is essential whenever the information being collected is of a subjective nature and is likely to be influenced by the wording or context of the questions themselves. Examples of such indicators are student and teacher attitudes and expectations, instructional methods, school organisation, and teacher assessments of principal leadership. This option has

Table 6.1 **Three options for collecting data for international education indicators**

Method of data collection	Type of data	Sample indicator	Advantages of method
COMMON INSTRUMENT	Knowledge, attitudes, processes, evaluation	Student achievement scores	Standard questions, ease of comparisons
COMMON REPORTING STANDARDS	Students, staff, finance, demographics, institution	Number of teachers by subject of assignment	Use existing data/systems, relate to within-country analyses
DESCRIPTIVE INFORMATION	Policies, curricula, programme innovations	Requirements for school completion	More specifics, context for comparisons

BEST COPY AVAILABLE

typically entailed using the same survey instrument in all participating countries, but it would also be possible to embed the questions or sets of questions into each country's own survey instrument.

The second option is to establish common standards for data reporting among countries. In the United States, many of the national educational statistics, as well as health and labour statistics, are based on state and national co-operation in data collection. The "Common Core of Data" on elementary and secondary education in the United States is an example. States collect these data as part of a system of "administrative records", a term typically used for data collection involving characteristics of individuals or institutions. Administrative records provide comparable data on objective characteristics that are not influenced by the context in which the data is collected. For example, states collect data on student enrolment ("membership") by grade from each local school district and then report the state totals to the National Center for Education Statistics (NCES). The aggregated data from states become the national education statistics. Comparability is ensured through common definitions of variables or data elements and common categories for aggregating and reporting the data. NCES is responsible for developing the definitions and the forms to be used in the system and for gaining states' approval and co-operation. Through state-federal co-operation, consistency and comparability across states can be established while allowing states the discretion to add other data collection that meets their needs.

The third option is a standard method of cross-national reporting and analysis used by the OECD, the United Nations, and other international organisations, as well as for cross-state analysis in the United States. It is used, for example, to compare laws and policies in a specific area, such as teacher licensing. It does not provide for comparable statistical indicators, but readers can infer differences and determine trends among the participating countries or states. This information is most useful when prepared in a common format to provide important contextual information for statistical indicators.

Commitment, cost and follow-through

A second issue in developing education indicators is ensuring commitment from decision-makers to follow through with implementation, and here costs are important. They will be high if the indicators require substantial collection of new data. In the United States, the implementation of state-by-state indicators first made use of indicators that largely drew on data already collected in many states or available from national surveys. This step made it possible to put some parts of the plan into effect immediately, and thus the effort was not viewed simply as a long-range goal.

The CCSSO model had three criteria for evaluating and selecting indicators:

- importance/usefulness of the desired indicator;
- technical quality of the data;
- feasibility of collecting and reporting the data on a state-by-state basis.

The criteria were applied in the order listed. Feasibility was placed last to avoid creating a set of indicators based only on currently available data. However, feasibility is important, particularly for some of the desired indicators.

Regardless of the means of collecting the data – using existing data, designing a new survey, or modifying existing data collection – strong commitment by a top decision-maker makes the task of the technical staff easier. Resolving differences in definitions and procedures or designing a new collection method are also easier if need, rationale, and funding have already been decided. Technical staff from participating education agencies are likely to be charged with determining whether data are available and meet desired standards of quality and comparability. If they are not available, these representatives work to determine the best means of designing a measure and collecting the required data. The technical staff's strong commitment to the indicators effort will improve the chances of producing accurate, useful data.

Analysis and interpretation of data

For international education indicators, appropriate comparative analysis and interpretation of education data are also critical. As an example, most countries collect data and report statistics on the number of students who complete secondary education. An indicator of secondary school completion is a likely candidate for cross-country comparisons. Three kinds of variations may present problems for analysis and interpretation:

- number of years or credits required for completion;
- curriculum content or type;
- societal value of secondary school education.

A common definition for data reporting on school completion may not succeed in incorporating all of these differences. Then, a planning group would have several options. It could decide that variation within a category does not invalidate comparisons (*e.g.* differences in the value of completing secondary education). The reporting definition could specify the acceptable amount of variation (*e.g.* completion might be specified as 11 to 13 years of elementary and secondary education). The reporting categories could be disaggregated according to possible sources of variation (*e.g.* each country reports the number of completers with a vocational curriculum, the number with a college preparatory curriculum, the number passing an equivalency test, etc.).

Differences in the "unit of analysis" are another possible source of non-comparability for data analysis and interpretation. The unit, or level, at which data are collected and aggregated may differ among countries. One country may have school completion data on each individual student, another may collect and file the data by school, and another by district or city. The unit in which the data are collected, filed, and available for reporting affects how the indicator can be reported and used at the international level. If the desired indicator is the number of teachers with certification in their teaching field, and if data are not collected and filed by individual teacher but by school or district, it would not be possible to develop an aggregate figure for teachers in a given field. Or, if data on student race/ethnicity are collected by district, a desired indicator of school variation in student racial/ethnic integration could not be produced.

To address these potential sources of non-comparability, the process of constructing an indicator should go beyond defining data categories and designing a reporting format; it should also provide a mechanism that gives each country opportunities for review and feedback both before and after data are reported. As a group, country representatives will have to decide how to minimise misinterpretation of data and determine the degree to which different interpretations can be tolerated. It is also important for a central staff to obtain and compare detailed information on each country's data collection procedures and system for filing the collected data.

Data procedures and reporting standards

Differences in methods and level of data collection and in definitions and wording of questions can be a source of difficulty for international comparisons. For example, each country participating in the assessments conducted by the International Association for the Evaluation of Education Achievement (IEA) uses a set of standard criteria to select a sample of schools and students. Differences in defining a school or an eligible student can produce biases in the country results. An IEA steering committee decides whether procedures have met a standard that allows for valid comparisons.

In the United States, differences among states in the definition of an elementary or secondary school teacher have produced problems for comparing teacher counts in the National Center for Education Statistics (NCES) annual report on education staffing. NCES has been working to set standard definitions and reporting categories for education staff. For international indicators, countries would need either to use common data collection procedures or to establish common standards for reporting data collected by individual country data systems.

Strategies for developing a set of education indicators

Several general strategies can be used with governments to move from the selection of indicators to the collection and reporting of valid, comparable data on the indicators. These strategies, based on experience in working with representatives of different education agencies to create education indicators, are key to implementing an indicator system.

A consensus approach to solving problems

A consensus approach to decision-making should be a basic strategy for developing international indicators. This involves joint decision-making among countries, and the decision may often be the best compromise position. Each country may have to adapt its methods, definitions, or procedures to meet the adopted standard. To work toward consensus, representatives need to adopt a ''problem-solving'' rather than a ''barrier-raising'' attitude. Governments must commit themselves to having aspects of their education systems – students, teachers, schools, and resources – compared with those of other countries. This may mean that some traditional ways of collecting and reporting data will have to be revised, that new data may have to be collected, or that ways of categorising and reporting data will have to be changed.

Ensure support by demonstrating the uses of indicators

Another strategy is to show policy-makers, administrators, and educators how the selected indicators will contribute to policy and programme decisions. The process of selecting indicators may not provide sufficient information for each country's education officials to plan for the kinds of data that would need to be generated and reported. Some may agree with the general idea of comparative indicators based on a specific indicator of interest to them, such as student achievement, without fully appreciating the range of indicators and data that would be produced. Others may not fully understand how education indicators can be used to track progress in one country through comparisons with others.

It is also important for managers and technical staff who are given responsibility for collecting and reporting data to understand the need for the indicators selected. For them, efforts to collect and report data should be viewed as serving informational and assessment needs within the country, as well as providing the basis for international comparisons.

One way of demonstrating use is to frame each indicator as a response to a policy or programmatic question. For example, an indicator of enrolments in advanced- versus elementary-level courses in a given subject can be viewed in light of the question, "What effect have changes in policy on school curriculum had on instruction?" Or, an indicator of per pupil expenditures on education can be viewed as a response to the question, "What effect does the degree of centralisation of finance of education have on equalising funding by local district or school?"

The uses of indicators can be explained to policy-makers, educators, and technical staff through workshops or seminars. An introduction to the indicators plan should be based on how the indicators will be useful for addressing the policy and programmatic questions being asked within a country. It should also emphasize the long-term building of an indicator system over the generation of data analyses to satisfy short-term information needs.

Project management and country representation

A final strategy concerns the management and operation of the indicators system. A central co-ordinating staff will be needed to provide day-to-day management of the system, including sending out forms, follow-up, data checking and editing, building and managing data files, analysis, and drafting reports. Senior staff will need to have expertise in education statistics, survey design and data collection procedures, computer database management, and statistical analysis. Representatives of participating countries should have a major role in system design, planning, review, and report development. The development of indicators is likely to require contributions of staff time from each country to aggregate and report the requested data. In addition, a central staff, offices, and equipment will need to be funded to ensure that the indicators system is developed, gains stability, and continues.

Staff contact persons will be needed in each participating education agency to provide continuity of indicators development over time. The contacts should be designated by a high government official who is authorised to commit staff time and resources to the indicators project. More than one type of specialist should be designated from each agency, including managers of information systems, staff assessment or research specialists, and programme or curriculum specialists. The contact persons in each country should include managers responsible for collecting and reporting data as well as staff responsible for using the indicators that are produced. The entire group of specialists would also have a critical role in setting and maintaining standards for data quality in collection, editing, and reporting.

Twelve steps in constructing indicators

Three basic options were outlined above for collecting and reporting data on international education indicators. It is likely that options one and three – common instrument and descriptive information – would be used to provide data for some of the desired indicators. Option two – common reporting standards – can be used to develop valid comparative indicators, but it would have to overcome special problems. Up to this time, it has probably been used less than the other two options for international education comparisons. The development of state-by-state indicators in the United States provides useful lessons for application of this option internationally.

Maintain commitment of decision-makers

At the selection stage, the initial commitment to state indicators was reinforced through communication with top officials at key points in development, such as designing data collection, developing reporting standards, and aggregating and reporting data. In some cases, steps were implemented some time after the initial commitment was made, and new officials had to be briefed and asked for commitment of their agency. This implies a need for staff continuity.

Establish country representatives

One way to reinforce commitment after the indicators are selected is to name staff representatives to work on constructing the indicators system. The state indicators network includes both data managers and data users from each participating state. Then, if revised or new data are required, a person responsible for design acts as the resource person. A letter was sent to state superintendents which gave criteria for selecting the representatives.

Conduct survey of countries concerning recommended indicators

Detailed information is needed about the availability of data and about the definitions, procedures and characteristics of each agency's data system and individual data elements. In the United States, therefore, a survey was conducted of the state network representatives, which allowed data managers and staff specialists to provide information about current data and about plans to expand their systems. The representatives worked as a team to respond to the survey and called upon other specialists where necessary. Information about all the indicators being considered was obtained through the same survey. For example, the survey included questions about state data on student demographics, teachers, school process, finance, and outcomes. Responses provided a breakdown of data available on each related indicator for 50 states. Follow-up questions were asked to obtain details on collection procedures and definitions.

Central tabulation and analysis

A central staff located at the Council of Chief State School Officers carried out the survey of states. Its role was important for survey design, answering questions, follow-up, checking and tabulating responses, and analysing results. A report was made on data availability and state data definitions and procedures for each of the recommended indicators.

Convene a task force of representatives to set priorities

The results of the survey provided sufficient detail about the recommended indicators to move ahead. Because the set of desired indicators exceeded the resources and capacity of states to actually collect and report data, it was necessary to set priorities. A task force comprised of state representatives from the network and experts on education indicators analysed the survey results and decided which indicators should be given highest priority for initial collection and reporting. Three criteria were used to evaluate and prioritise indicators: importance/usefulness, technical quality of data, and feasibility. The survey of state data availability helped determine feasibility (how many states already collected the data needed for an indicator) and data quality (methods used for data collection, level at which data were collected and aggregated). The survey also provided state responses concerning importance/usefulness of recommended indicators.

The three criteria for setting priorities need to be considered together. Indicators with the highest ratings on importance/usefulness may not be the most feasible. For example, a highly feasible indicator of school process might be a country's policy on time for teaching core academic subjects in elementary and secondary schools. However, this indicator may be rated low on importance and usefulness if representatives find that time allocations are a poor measure of time actually spent in the classroom. An indicator of the time teachers actually spend teaching specific curriculum content might be highly rated on importance/usefulness but be rated low on quality and feasibility.

Using this approach, a task force could produce an evaluation of the recommended indicators on the basis of the existing and planned data systems of the participating countries and of practical judgements on which indicators will be most informative and useful for decision-makers, planners, and education analysts. A list of priorities for the initial list of recommended indicators can provide a plan for beginning work on collecting and reporting data.

Design a common reporting system

When each education agency's capacity to produce data for each priority indicator is known, it is possible to design a plan for reporting data using a common set of definitions and categories. Copies of each country's data collection forms and definitions are needed, and in some cases, follow-up questions are needed to clarify data collection procedures, definitions, and methods of data aggregation.

Several options were considered in designing the data reporting format for education indicators in the United States. One was to report only aggregate state totals on each requested element and category. For example, for an indicator of student enrolment in core curriculum subjects, the total enrolment would be reported for each of several common categories in each subject by grade level. (In science, one category might be "total enrolment in first-year biology, grades 9-12".) A second option was to have the indicator include total state enrolments and totals disaggregated by student demographic groups, such as gender or race/ethnicity. A third option was to have each state report data disaggregated by levels within the education system, such as school, district, city, county, or, possibly, individual student or teacher. With the third option, states would have reported data by computer tape or disk. The second and third options presented the advantage that analyses could include a wider range of comparisons than simple state totals. With each option, it was anticipated that states would report numbers and then rates would be computed for purposes of state-to-state comparison.

BEST COPY AVAILABLE

Construct data categories

Several objectives should be considered when developing data categories for the desired indicators. One is to construct broad categories that will maximise the numbers of countries that can report while ensuring that data are comparable from country to country. In the development of state indicators, it was found that half the states collected data on the subject assignments of teachers by their "primary" versus their "secondary" subject, while the others collected data on assignments by teaching period. The reporting plan defined categories for teacher assignments that included both methods.

A second objective in constructing data categories is to maximise the usefulness of comparative data by obtaining input from primary data users. Many states expressed interest in tracking the effects of changes in state graduation requirements on the types of courses students take in science and mathematics. Accordingly, the high school course enrolment categories for science and mathematics were designed to differentiate courses according to four levels: basic or applied, general, advanced, and advanced placement. Categorising student enrolments by these levels provided valid comparisons across states as well as a method of tracking effects of changes in requirements.

A third objective is to design data categories and definitions that are useful to countries that will be developing or revising data collection forms. Ideally, the definitions and categories should anticipate the data needs of decision-makers and educators. The design should not be too complex and should be applicable to the needs of most education systems. Thus, it should be modelled neither after the country with the most detail, nor after the one with the fewest categories and most simple definitions.

In the United States, an initial draft of a common reporting plan and data categories was prepared by state indicators project staff. However, the final design, categories, and definitions were decided by representatives from the states. These representatives ensured that the design would reflect the states' policy and programmatic needs. The indicators system will not be useful unless it reflects the participating governments' needs and interests.

Review reporting plan with each country

A mechanism should be established to allow each country's representatives to review and analyse the implications of the data reporting plan. The planning team should provide all participating countries with an opportunity to make comments, suggest revisions, and discuss the data reporting plan with the planning team.

For state indicators, a series of regional workshops was organised for state representatives. The workshops included a full explanation of the data reporting plan, small group discussions of the uses of the selected indicators, and meetings with each state's representatives. The advantages of workshops for direct review and discussion over a review through the mail, include: immediate discussion of any questions or problems, sharing and exchange of ideas about use of indicators, and opportunity to develop a timetable and plan for each state's participation.

Conduct pilot study

Using data from a sample of eight participating states, a pilot study was conducted to test the reporting plan. The draft reporting plan's definitions and categories were used to identify errors or gaps in definitions, missing categories, misinterpretation of instructions, and problems of non-comparability. With the pilot study results, trial comparisons were made among states. A task force of state representatives reviewed the results and suggested ways of correcting problems in the reporting plan.

Prepare and send reporting instructions and forms for initial data reporting

The reporting form for state indicators included a match of state data codes with the common reporting categories for each indicator. The match was based on state data collection instruments and code books prepared by central project staff and reviewed by state specialists. The data on state indicators were reported as referring to one date in the school year (October). The instructions and forms were mailed to the states in early September. Then, reminders were sent one month prior to the date the data were due. Instructions were sufficiently detailed and comprehensive to allow a data specialist or computer programmer to generate the requested data without prior knowledge of the project.

Edit data and prepare initial report

The central staff for state indicators conduct data quality checks and corresponds with contact persons in each state to complete data editing and revisions. The standards for reporting were set by the planning task force, but the central staff are responsible for quality control of each state's data report. The initial drafts of summary analyses and tables are shared with the state representatives prior to any detailed analyses. The drafts are used to double-check each state's data as they are listed alongside data of other states.

Decisions about analyses using the edited and corrected data should be based on the conceptual framework for the indicators and on discussions by the planning team. Since state-by-state analyses produce a large number of statistics, the types and breadth of analyses are carefully considered. For many indicators, bivariate analyses are useful, such as number of teachers in a field by age and state. With the release of an initial report on state-by-state indicators, state indicators are likely to provide strong motivation for improving the data capacity of those countries that cannot deliver the data requested by the OECD.

Establish cycle of data reporting and expanding indicators

The plan for reporting on international indicators should specify the expected frequency with which data will be reported on the high-priority initial indicators as well as the schedule for expanding the system of indicators. States had more incentive for collecting, editing, and reporting quality data if periodic reporting was expected. When the initial report was released, the next expected reporting period was publicised. The report also outlined steps and time periods for implementing the remaining indicators in the model. The published plan for indicators should specify when and how other indicators are expected to be added to the reporting cycle.

Conclusion

This chapter addresses the question of constructing a system of international education indicators. It does not address the issues of the need for these indicators or the kinds that should be developed. Assuming that a set of indicators is selected and that a number of countries have agreed to participate, the information presented here may be useful to those who develop a system of indicators.

To develop a valid and useful system of indicators, several issues will have to be resolved. A major problem is obtaining and reporting data that will be comparable from country to country. Methodological approaches to constructing indicators can vary, even from one indicator to another. A substantial part of the work of constructing indicators will be the development of a consensus on specifications for data collection, including definitions, categories, and reporting standards. Another issue is resolving differences in interpretation of data and determining the appropriate unit of analysis for international comparisons. Countries will need to maintain their commitment to the indicators system over time, allocate resources to the project, and provide technical staff that will follow through on the initial policy decisions to develop indicators.

Three broad strategies, or approaches, may increase the chances for effective development of international indicators. Country representatives will need to develop a problem-solving, consensus approach to the project, so that participants will tend to seek solutions and compromises rather than raise barriers or find reasons to curtail a country's involvement. Top education administrators from each country should select managers and technical specialists who will provide continuing representation and contact for indicators development. In the planning and development phases, leaders and staff should prepare information and materials that demonstrate the usefulness of international comparable indicators for decision-makers, administrators, educators, and researchers.

Specific steps for constructing education indicators based on data collected by each country and reported through a common set of standards are proposed. They include surveying and analysing existing data systems, reaching consensus on priorities for indicator development, designing a data reporting plan and specifying definitions, reaching agreement on a schedule for data collection and reporting, and evaluating the comparability and quality of data. These steps give a picture of the kinds of plans, decisions, group work, and staff contributions that are likely to be necessary. They are based on experience with states and would have to be adapted for developing international indicators.

The recent experience of the OECD indicators project shows that many countries are committed to education indicators. Putting such indicators in place will require resources, co-operation, and commitment. This description of some of the problems, issues, strategies, and steps in constructing indicators may assure the transition from commitment to realisation.

Bibliography

BLANK, R.K. (1989), "Development of a 50-State System of Education Indicators: Issues of Design, Implementation and Use", paper read at the annual meeting of the American Educational Research Association, Council of Chief State School Officers (CCSSO), Washington, D.C.

BLANK, R.K. and DALKILIC, M. (1990), *State Indicators of Science and Mathematics Education*, CCSSO, Washington, D.C.

BOTTANI, N. and WALBERG, H.J. (1992), "What are Educational Indicators for?", in *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, pp. 7-12.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS (1985), "Assessment and Evaluation Plan of the Committee on Co-ordinating Education, Information and Research", CCSSO, Washington, D.C. (mimeo).

COUNCIL OF CHIEF STATE SCHOOL OFFICERS (1989), *State Education Indicators*, CCSSO, Washington, D.C.

SELDEN, R.W. (1990), "Developing Educational Indicators: A State-National Perspective", *International Journal of Educational Research*, Vol. 16, pp. 383-393.

Chapter 7

The Importance of Opinion in Studying the Education System

by

Robert Ballion
Ministry of Education, France

The end product of the education system (a structure for helping people develop a range of different abilities) is largely dependent on those who are involved in the process.

The actors in the system

Professionals whose responsibility it is to provide the service operate within a framework which is physical, governed by rules, and, due to the image and norms it projects, symbolic, and which influences, and even prescribes, their conduct. Nonetheless, these constraints leave them a considerable degree of autonomy, not only because education professionals have been given a mandate to interpret their job for themselves within the framework provided by programmes, directives and rules, but also because teaching is perceived as being more than an occupation, which implies a degree of personal involvement and thus a certain freedom.

As for those who receive the service, pupils and students, the part they play is even greater, because their development – intellectual, emotional and moral – is ultimately dependent on what they are willing to invest in the process.

What the various actors do is influenced both by the structure in which they operate and by the attitudes they bring with them. These are more or less stable and deeply rooted attitudes, and they shape the response of individuals to the educational structure.

In order to analyse the educational process, it is useful to identify these attitudes for two reasons: first, because attitudes are an element of the learning process and affect educational outcomes; and second, because they shed light on the conditions under which the educational process takes place.

Attitudes cannot be measured directly. They have to be inferred either by hypothesing that observed behaviour is only meaningful in the light of underlying attitudes, or, more directly, by examining the views that people themselves express. When people formulate judgements, evaluate, express their needs, beliefs, or values, they give evidence of a scale of values which sheds light on their actions.

Surveys of opinion

Of the various methods used to gather opinions, the questionnaire has proved to be the most often adopted and the most systematically used for studies which take account of the point of view of the actors. Opinions take their place in this way among other explanatory models of the patterns of variables.

Outside the scientific domain, opinion gathering plays an increasingly important role in our society through polls, which take a sample of the population and try to elucidate a problem or a specific situation by analysing the reactions of different members of the public. The sample may cover the entire population of a (national or local) geographical area, or certain categories of the population, *e.g.* parents of school pupils, for questions of education.

In the case of the education system, the existence of these polls offers the advantage of constituting a large body of already available and continuously updated material. By analysing their content and origin, much can be learned about the questions a society thinks it important to ask about its education system and how the concerns differ depending on the source of the opinion poll (public authorities, the press, unions, parent associations, economic actors).

Opinion polls provide information. They make it possible to understand how well the education system is perceived to be working and to make choices among the many different points of view. By its requests for polls, a society expresses what it sees as most important, or at least where its priorities lie. Among the questions asked by the polls, some can be chosen as indicators that can allow a given society to understand the views of its members on education and on the system that provides it, and to comprehend the evolution of these views over time through the construction of series.

If a number of countries have common indicators, many useful lessons can be learned. Each country can compare its school system to that of others and evaluate its situation in that light. For example, 40 per cent of French teachers say that they would "like to leave teaching to take up another job" (IPSOS, 1985). It would be interesting to know whether this indicator of "disenchantment" or "disillusionment" in France is very different from that of other countries whose society and education systems are comparable.

The existence of indicators for international opinion would also make it possible to form views about the factors that influence outcomes. Disparities in outcomes could be compared to the disparities in opinions that might result in behaviour that affects outcomes.

Opinion polls on education in France

The Directorate of Information and Communication of the Ministry of Education reviewed education polls conducted in France between 1973 and 1988. Over those 15 years, there were 349 polls, which were not evenly distributed but became more numerous as time went by. There were ten polls in 1973/74 and 79 in 1987/88. The pace began accelerating at the beginning of the 1980s.

It is difficult to describe the polls by their content because, for the most part, they dealt with a number of themes. Nonetheless, the Directorate of Information and Communication used a scheme which classifies them according to both the dominant theme and the population that was polled.

However imperfect, this classification reveals that most attention is paid to those who are being educated, the young people who are the object of 145 polls out of 349; the views of economic actors are instead only rarely represented, since only eight polls concerned firms (see Table 7.1).

The breakdown of these polls by topic and source makes it possible to evaluate the relative weight of public authorities and public opinion in relation to the demand for information about the education system. It also makes it possible to judge where their dominant interests lie.

Table 7.2 shows that the Ministry of Education requested only a small number of the polls, less than 8 per cent. Even when those initiated by other ministries and organisations in the public sector such as teachers' unions and parents' associations are added, they only account for a quarter of the total number. Most (71 per cent) are carried out by the press, both professional and lay, and most concern pupils and students.

The fact that the Ministry of Education is so poorly represented among those commissioning polls does not mean that it attaches no importance to collecting information about the education system, but simply that it does not give high priority to this method. The public administration has two main channels for monitoring the education system: the General Inspectorate (*Inspections générales*) and the Directorate for Evaluation and Planning (*Direction de l'évaluation et de la prospective*). There are also the studies of agencies such as the National Institute of Educational Research (*Institut national de la recherche pédagogique*) and the Centre for Study and Research on Employment Qualifications (*Centre d'étude et de recherche sur les qualifications et l'emploi*) or those carried out by the research community. As a result, the public sector has means of obtaining information which allow it to dispense with polls.

The medias' interest in polls on education may be seen as a result of a growing concern on the part of the public to evaluate its school system and its results. The public's

Table 7.1 **Overview of education polls conducted in France from 1973 to 1988**

Orientation	Number	Per cent
Polls of young people	145	42
Polls of teachers	23	7
Polls of parents	37	11
Polls of industry and commerce	8	2
Polls treating various educational problems, essentially relationships between public and private sectors, school year and school cycle, and the introduction of new information technology	49	14
Polls of the general public dealing with education as a whole	87	25

Source: French Ministry of National Education (no date)

BEST COPY AVAILABLE

Table 7.2 **Agencies that initiated education polls in France, and their principal target groups (polls 1988-1990)**

	Target groups						
	Students	Teachers	Parents	Industry	Specific problems	Public opinion	Total
Ministry of Education	3	1	1			1	6
Other public and quasi-public agencies	7	1	3	1		2	14
Specialist educational press	13						13
General press	26	1	4	1	1	10	43
Other agencies	2				1		3
Total							
(n)	51	3	8	2	2	13	79
(%)	65	4	10	3	3	17	100

Source: French Ministry of National Education (no date).

principal interest is with pupils and students, and it appears to be dissatisfied with its access to educational information, since it feels the need to produce such information itself.

Some recurring themes

Even without a systematic analysis of the content of polls, some recurring themes can be easily discerned. They seem to indicate the areas considered important by public opinion and thus to reveal the concerns that the system should address.

In one form or another, views of the education system are sought, for example by comparing schools to other social institutions. Thus, in a comparison of confidence rates ("Do you have more confidence in ...?"), schools rank first (84 per cent), followed by the police (70 per cent), enterprises (70 per cent), and banks (64 per cent). Political parties come last with 15 per cent and are preceded by trade unions (32 per cent) and the media (42 per cent) (SOFRES, 1989).

Overall, 63 per cent of the French say that they are satisfied with their school (very satisfied: 15 per cent; satisfied: 48 per cent). The parents of pupils are even more likely to be satisfied (85 per cent) than people who do not have children in the education system (53 per cent) (Publimétrie, 1987).

Perceptions of the role and purposes of schooling can be categorised into four groups. First, by a wide margin, comes preparation for a career, while development of intelligence and acquisition of knowledge as well as the forming of personality have a place in the middle, and preparing young people to become good citizens trails far behind. When the French are asked to state their priorities for secondary school education, they rank the goals as in Table 7.3.

Table 7.3 **Public attitudes towards goals for secondary education**

Goal orientation	Per cent assigning priority
Preparing for a job	47
Preparing for a changed world	40
Developing character	37
Acquisition of general culture	36
Developing intelligence	22
Developing creativity	13
Developing citizenship	4

Source: CREDOC (1988).

When parents alone are asked to judge the teachers of their children, 83 per cent say they are satisfied, but when asked to suggest what should be done to improve teaching they offer a considerable range of replies. When both parents and students are polled on the criteria of a "good school", the former emphasize discipline and the latter the need for high-quality teaching. They are in agreement on the importance of criteria such as success in examinations, good relationships between pupils and teachers, and an environment free from violence and drugs. Both parents and *lycée* pupils have a positive view of their school: 66 per cent of parents and 57 per cent of pupils think that it is good; 33 per cent and 41 per cent, respectively, consider it average; and only 1 per cent of parents and 2 per cent of pupils think that it is poor (Phosphore, 1988).

Questions addressed specifically to pupils deal with their school life, what they expect from their teachers, and the opinions they have about their role in school administration. Pupils expect principally that the school should prepare them for a career. In this respect it is significant to note that their view is at odds with that of their teachers. Thus, when asked to choose among two primary objectives, 61 per cent of *lycée* pupils, against 28 per cent of their teachers, saw preparation for a career as the principal goal of education, whereas 67 per cent of teachers, against 39 per cent of *lycée* pupils thought that education should principally form personality. Among the teachers, 5 per cent had no response to this question (SOFRES, 1989).

The emphasis placed on professional goals explains why 97 per cent of the 15-18 year age group think that "the development of relationships between school and businesses is a good thing" (WSA, 1987). Young people are primarily concerned with their career, and 75 per cent of them think that this preparation must take place in the school (Publimétrie, 1982).

Secondary school teaching does not seem to meet that expectation. Only about half of *lycée* pupils (54 per cent) respond positively to the following statement: "What I learn in the *lycée* interests me because it will be useful to me in the future" (CSA, 1991). The most common complaint about the education system from 15-24 year-olds is that it offers "a form of teaching which takes no account of changes in the real world" (Démoscopie, 1986). Only 18 per cent of secondary school pupils think that school prepares them properly for a career. The number rises to 53 per cent when "activity" replaces "career" (Publimétrie, 1982).

Teachers' views of their role differ from those of families and pupils. They place a high priority on the importance of developing critical faculties and imagination in young people (44 per cent) and on acquisition of sound general culture (32 per cent), while they relegate to last place what most concerns young people, preparation for a career (20 per cent) (CREDOC, 1988).

Three-quarters of teachers are satisfied with their job and "if they were to do it again", 65 per cent of them would take the same path. This is so, despite the fact that almost all feel that their salary is inadequate (92 per cent) and that they lack any real career prospects (77 per cent).

Two findings are worth highlighting. One is that teachers have a mistaken view of their public image. While 83 per cent of parents say that they are very or quite satisfied (27 and 56 per cent, respectively) with their children's teachers, and the children share their opinion, but almost all teachers (93 per cent) complain about the lack of social recognition of what they do.

The second point is that teaching is seen as an exciting, but very difficult profession, and that the training for it is unsatisfactory: 43 per cent of teachers say that their training was inadequate with respect to their subject matter, and 75 per cent with respect to pedagogical methods.

Opinion polls on education provide a wealth of information that is infrequently and poorly used because it is difficult to apply, given the range of contexts in which polls are conducted. Another reason is the lack of co-ordination. Polls are seen as providing specific evidence at a particular time and relative to that occasion, and this makes it difficult to standardize the process. Even questions on the same topic are put in different words and those on the workings of the system are not consistently taken up over time.

Conclusion

All of this makes it difficult to provide an overview or a general diagnosis of the education system from results available at a given time and to make comparisons over time for most of the questions addressed. This is unfortunate, because such methods of enquiry provide a considerable amount of useful information for the evaluation of the education system and for policy-making.

Polls make it possible to make appraisals on the basis of public opinion. It is the only way to bring to light expectations with respect to the education system and the attitudes that can influence the processes that affect the nature and quality of outcomes.

The elaboration of international indicators would make it possible to obtain standardized measures over time and in specified regions, and thus, to measure expectations of public opinion and its various components about the role to be played by the education system, the levels of satisfaction and confidence, and the nature and importance of the role played by different actors in the functioning of institutions. It would constitute an appreciable advance on the road to making phenomena in the area of education more objective.

Bibliography

CREDOC (1988), ''Sondage CREDOC'', ministère de l'Éducation nationale, August.

CSA (Conseils Sondages et Analyses) (1991), ''Sondage CSA'', *Phosphore*, Paris, September.

DÉMOSCOPIE (1986), ''Sondage Démoscopie'', *Lettre de l'Éducation*, Paris, November.

IPSOS (1985), ''Sondage IPSOS'', *Le Monde*, Paris, 10 September.

MINISTÈRE DE L'ÉDUCATION NATIONALE (no date), ''Inventaire des attitudes, sondages et travaux disponibles'', direction de l'Information et de la Communication, ministère de l'Éducation nationale, Paris.

PHOSPHORE (1988), ''Sondage *Phosphore* : Qu'est-ce qu'un bon lycée'', *Phosphore*, Paris, February.

PUBLIMÉTRIE (1982), ''Sondage Publimétrie'', Fédération des Conseils de parents d'élèves, Paris, April.

PUBLIMÉTRIE (1987), ''Sondage Publimétrie'', Fédération des Conseils de parents d'élèves, Paris, September.

SOFRES (1989), ''Sondage SOFRES'', Observatoire interrégional du politique, Paris, October.

WSA (1987), ''Sondage WSA'', *Phosphore*, Paris.

Chapter 8

Process Indicators for School Improvement

by

David Hopkins
University of Cambridge, Institute of Education, United Kingdom

This chapter deals with the relationships among school processes, system level characteristics and educational inputs, and their implications for the development of process indicators. It adopts a "school improvement" perspective. This results in a tension between, on the one hand, an "analytic" and "objective" description of the various indicators of "school process" that contribute to increased "quality" and, on the other, the distinctly "dynamic" character of much of the research-based knowledge of the school process.

The chapter is divided into three parts. The first defines the nature of process indicators and places the discussion in the wider context of research on school effectiveness, school improvement and policy analysis. It then addresses more fundamental questions about the nature of process and suggests a reconceptualisation. The summary contains a list of process indicators based on this initial review of the literature.

In the second part, conventional process indicators are reviewed in the light of the reconceptualisation. Although this entails revisiting familiar territory, the intention is to illuminate it from a different, and hopefully more enlightened, perspective. The research on discrete indicators identified in the first part is reviewed in the light of the complexities of school culture. Strategies for school improvement and case studies of external support initiatives, which view process indicators holistically, are also taken into consideration. Finally, a more refined set of process indicators is presented.

The third part of the chapter considers the application of this set of process indicators, particularly as a basis for international comparison, and its contribution to school improvement is assessed.

Definitions

Performance and process indicators

Indicators are the substantive focus of the INES project. Consequently, as phenomena, they are dealt with more directly and in more detail elsewhere in this volume.

However, it is necessary to define briefly the way in which the term is used here. A performance indicator is usually regarded as a quantitative measure for judging the performance of an individual, group, institution or system. It refers in the main to either inputs or outputs. In reference to schools, an input could be, for example, the measured cognitive ability or socio-economic status of a student on entry; an output could be a student's examination scores on leaving the school.

In recent years there has been considerable discussion of the uses of indicators in education; in the INES project, they are referred to as education rather than performance indicators. Many different groups have produced lists of indicators which are generally used for purposes of accountability and comparison. There has also been wide discussion of the problems of devising and using them. Their conceptual, epistemological and cultural bases are often heatedly debated, together with issues such as reliability, validity, fairness, credibility and justification.

Performance indicators in education generally operate at three different levels (national, local and school). They are a mixture of quantitative and qualitative measures focused on school processes, inputs and outputs. Process indicators are usually regarded as measures of the internal features of schools – such as organisational arrangements or the quality of teaching and learning – that link inputs to outcomes. They are the black box in the middle of the flow diagram that accounts for the "value added" to the student by the school. However, life in schools is rarely as simple as this.

In practice, these elements interact. At the national level, performance indicators are usually quantitative and relate to inputs and outputs. National inspectors, however, often use a combination of quantitative and qualitative measures and look at processes as well as inputs and outcomes. At the local level, performance indicators also reflect the criteria used in local inspections as well as specifications and standards related to a range of curriculum initiatives. At the school level, although performance indicators are used in all these ways, principals and teachers are increasingly using a school-generated performance indicator, often called a success criterion, as a means of planning, implementing and evaluating their own development. This indicator is distinctive in that it refers to future rather than past performance, relates to a planned target and suggests standards for evaluation (Hargreaves and Hopkins, 1991).

Despite these differences, the term "performance indicator" is often used for all these activities. It is little wonder, therefore, that confusion arises. Fortunately it is not the task of this chapter to sort out this problem, although it does aim at achieving some clarity on the nature of process indicators. At this stage, it is prudent to stay with the commonsense definition already advanced: process indicators are measures of those internal features of schools that translate student ability into student achievement. They concern what the school or educational system does to create quality. In the following section, the research literature most closely associated with educational process is reviewed, in an attempt to give more texture to this definition.

School effectiveness and improvement

A vast amount of evidence has been accumulated to support the commonsense notion that the internal features of individual schools can make a difference to pupils' progress. The research on "effective schools" consistently demonstrates correlations between student achievement on tests of basic skills and a stable set of school organisation and process characteristics. By way of contrast, the emphasis in school improvement studies has been on creating conditions within schools that enable them to become more effective. The connections between these two fields of study are consequently of some importance in attempting to define a generalisable set of process indicators for quality schooling.

Much of the early work on effective school correlates was done by Edmonds (1978), who lists:

- emphasis on student acquisition of basic skills;
- high expectations for students;
- strong administrative leadership;
- frequent monitoring of student progress;
- orderly climate conducive to learning.

Subsequent research on effective schools in the United Kingdom (Rutter *et al.*, 1979; Mortimore *et al.*, 1988) and the United States (Purkey and Smith, 1983) has found that similar internal conditions typically obtain in schools that achieve higher levels of outcomes for their students. Not only does the effective schools research conclude that schools do make a difference, but there is also agreement on two further issues. First, that differences in outcome are systematically related to variations in the school's climate, culture or ethos. Second, that the school's culture is amenable to alteration by concerted action on the part of the school staff. Although this is not an easy task, the evidence suggests that teachers and school staff have more control than they may have imagined over their ability to change an existing situation.

There is broad agreement that the following eight criteria are representative of the organisational factors characteristic of effective schools (Purkey and Smith, 1983):

- curriculum-focused school leadership;
- supportive climate within the school;
- emphasis on curriculum and teaching;
- clear goals and high expectations for students;
- a system for monitoring performance and achievement;
- ongoing staff development and in-service;
- parental involvement and support;
- Local Education Authority (LEA) and external support.

These factors do not, however, address the dynamics of schools as organisations. There appear to be four additional factors that stimulate school improvement. These so-called process factors provide the means of affecting the organisational factors and making the system more dynamic. They have been described as follows:

- A feel for the process of leadership – this is difficult to characterise because the complexity of factors involved works against rational planning. A useful analogy would be that organisations are to be sailed rather than driven.
- A guiding value system – this refers to a consensus on high expectations, explicit goals, clear rules, a genuine caring about individuals, etc.
- Intense interaction and communication – this refers to simultaneous support and pressure at both horizontal and vertical levels within the school.
- Collaborative planning and implementation – this needs to occur both within the school and externally, particularly in the local education authority (Fullan, 1985, p. 400).

BEST COPY AVAILABLE

It might appear that those working within the effective schools tradition have provided a sufficient list of process indicators. However, as a source of process indicators for school improvement, school effectiveness research raises a number of specific problems.

The first is a conceptual one. It is obvious from the three lists just cited that the correlates are often of a different order; "strong leadership", for example, is very different from "a school's guiding value system". Similarly, the distinction between "organisational" and "process" factors is not clear-cut, and climate and leadership appear under both headings. In any case, what do the terms mean in practice? "Emphasis on curriculum and teaching" is far from precise, and many of these terms are subject to widely differing interpretations. Much more conceptual work must be done on the effective school criteria before they can provide an unambiguous guide for action.

The second problem has to do with administrators who, in their search for simple solutions to complex problems, may depend naively on research evidence and test scores for solutions to their pressing educational concerns. As Cuban (1983) points out, too narrow an interpretation of school effectiveness criteria leads to increased standardization, a narrowing of the educational agenda, and complacency in schools that have good examination results. Cuban argues that the question should really be: "How can the broader, more complex and less easily measured goals of schooling be achieved as we improve test scores?" In this respect, the effectiveness criteria have too narrow a focus for a school improvement strategy.

The increased sophistication of recent school effectiveness studies raises the third problem. New analytical techniques have made possible more detailed investigation of the differential impact of school effectiveness on sub-groups. Mortimore *et al.* (1988) for example, found that there was some variability in progress in reading between boys and girls in the same junior schools. Nuttall *et al.* (1989), in a study of Inner London Education Authority (ILEA) secondary schools, found that the effectiveness of a school varies along several dimensions and over time. These findings suggest that the existing criteria lack the comprehensiveness required for a full school improvement strategy.

A collection of school effectiveness case studies conducted in the United States (Taylor, 1990) reveals the fourth problem. Although the studies showed clarity and consensus about the effective school correlates, the nature of the process that leads to effectiveness was little discussed. Nowhere was the process of translating the correlates into a programme of action sufficiently articulated.

Finally, the criteria for school effectiveness tend to be treated individually. Although this is not always the case, the research design in most of these studies results in a list of individual factors rather than a holistic picture of school culture. When school "ethos" or culture is discussed, it is usually the result of *ex post facto* conceptualisation.

At best, the effective schools criteria can provide a starting point for identifying some of the individual process indicators for school improvement. School improvement research may offer another avenue of approach, as it is more oriented towards action and development. It embodies the long-term goal of the "problem-solving" or "thinking" or "relatively autonomous" school, to be achieved by developing strategies that strengthen the school's organisation, as well by as implementing curriculum reform.

This approach is exemplified in the work of the International School Improvement Project (ISIP) sponsored by the OECD and the knowledge gained from it (Hopkins, 1987). School improvement was defined in ISIP as:

> "... a systematic, sustained effort aimed at change in learning conditions and other related internal conditions in one or more schools, with the ultimate aim of accomplishing educational goals more effectively." (van Velzen *et al.*, 1985, p. 48)

This obviously implies a very different way of thinking about change from the "top-down" approach so popular with policy-makers. When the school is regarded as the "centre" of change, strategies for change need to take this perspective into account. The ISIP served to popularise the school improvement approach to educational change, which rests on a number of assumptions:

- *The school as the centre of change.* This means that external reforms need to be sensitive to the situation in individual schools, rather than assuming that all schools are the same. It also implies that school improvement efforts need to adopt a "classroom-exceeding" perspective, without ignoring the classroom.
- *A systematic approach to change.* School improvement is a carefully planned and managed process that takes place over a period of several years.
- *A key focus for change is the "internal conditions" of schools.* These include not only the teaching-learning activities used in the school, but also the schools' procedures, role allocation, and resource use that support the teaching/learning process.
- *Accomplishing educational goals more effectively.* Generally speaking, the school's purpose is to fulfil educational goals for its students and society. This suggests a broader definition of outcome than student scores on achievement tests, however important they may be. Schools also serve the more general developmental needs of students, the professional development of teachers, and the needs of its community.

ISIP was a decentralised project divided into the five major areas of study for school improvement: school-based review; the role of school leaders; external support; research and evaluation; and policy development and implementation. Each working group was concerned to examine current provision and to develop strategies for school improvement policy and practice. This was both ISIP's strength and its weakness. The structure allowed for researching and developing individual strategies in some depth but militated against holistic and integrated approaches. This seems to be a fairly general problem in most schemes for school improvement.

So far, the discussion has centred on general strategic approaches to change. One other area of school improvement research should be briefly mentioned, specific individual approaches. US research on this topic relates specific individual strategies to student test score as a measure of effectiveness. Using this approach, in inevitably limited situations, some school improvement strategies have produced startling results. For example, the research on the application of different models of teaching, following the example of Joyce and Weil (1986), has resulted in consistently higher test scores in some classrooms. Co-operative learning is also achieving consistent improvements in many classrooms across the United States. The direct use of educational technology in classrooms and collaborative approaches to staff development such as coaching have resulted in positive effects on student learning.

Unfortunately, not all these strategies are effective all the time and in every setting. When they are not, it is often because they have ignored the school culture. Many school improvement strategies implicitly assume that behind the "door" is a network of interconnecting pathways that lead inexorably to school improvement. This is not so. Too often they focus on individual change, discrete projects, and individual teachers and classrooms, rather than on how these changes fit in with or can be adapted to the organisation and ethos of the school.

In the quest for process indicators, studies on school effectiveness and improvement help to define the territory, but they also serve to demonstrate its complexity. As a consequence, the "black box" notion must be rejected as far too simplistic, and the issue of "process" must be confronted directly.

What is process?

Once again, earlier work is a good place to start. Although there are many studies that focus on process indirectly, few papers take it as their main focus. Of those that do, those of Scheerens (1990) and Oakes (1989) are of particular interest.

Scheerens begins by stating that "a context-input-process-output model is the best analytic scheme to systemise thinking on indicator systems" (p. 62). Although apparently convinced of the importance of process indicators in understanding what goes on in schools, Scheerens doubts that they can be "used as a basis for judging the performance of an educational system" (p. 63). He warns of the dangers of sweeping evaluative conclusions and argues that process indicators should always be linked to output indicators: "process indicators then have the function of offering hypothetical explanations of why certain schools, or school systems, do better than others" (p. 63).

Given his commitment to the input-output model and his demonstration that no "empirically supported causal models of educational performance, from which the importance of specific process measures could be deduced" (p. 63) exist, it is not surprising that Scheerens turns to the "process-product" research on school effectiveness to derive his own specification of process indicators. ("Process-product" research designs provide the basis for most school and classroom effectiveness studies; they work backwards from school or classroom outcome measures to establish correlations between these scores and putative school or classroom processes.) Working within this tradition, Scheerens posits a set of process indicators at the school and classroom levels (p. 73):

School level:

- degree of achievement-oriented policy;
- educational leadership;
- consensus, co-operative planning of teachers;
- quality of school curricula in terms of content covered and formal structure;
- orderly atmosphere.

Classroom level:

- time on task (including homework);
- structured teaching;
- opportunity to learn;
- high expectations of pupils' progress;
- degree of evaluation and monitoring of pupil's progress;
- reinforcement.

Scheerens (1992) claims that this list of process variables are the "most relevant in exploring the causes of achievement differences between schools" (p. 73). He supports this claim by demonstrating a high degree of consistency between his list and other lists of process indicators. When compared with the empirical research, however, there is a higher degree of support for the classroom than school level characteristics.

Although its specification of individual process characteristics is very helpful, this approach presents problems when viewed from a school improvement perspective. The commitment to a particular form of empiricism inevitably leads to a "black box" view of process; also, the role of context, although acknowledged, is underplayed.

Oakes, in contrast, gives full weight to the role of school context in her review of process indicators. Taking a policy perspective, she argues that school evaluations that focus on only a small range of test score indicators encourage teachers to narrow their programmes in order to "look good" for the limited range of criteria. She argues that these measures should be balanced with "equally influential indicators of valued school characteristics". Her argument for context is as follows:

> "Context indicators, then, provide information about central features of the educational system. We must monitor and observe these features to learn more about how the system works and the conditions under which particular student experiences and results take place. If policy makers choose not to monitor context, they will fail to recognise that school characteristics mediate the effects of educational inputs (for example, resources and state and local district policies). They will also ignore how school characteristics can influence the classroom interactions that affect learning. Doing so, they will create monitoring systems that provide a superficial and simplistic portrayal of the educational system – one based entirely on results." (1989, p. 183)

Based on a comprehensive review of the literature, Oakes suggests three general constructs that can serve as a basis for developing school context indicators: access to knowledge, press for achievement, and professional teaching conditions.

Access to knowledge:

- teacher qualifications;
- instructional time;
- course offerings;
- class grouping practices;
- materials, laboratories, equipment;
- academic support programmes;
- enrichment activities;
- parent involvement;
- staff development;
- faculty beliefs.

Press for achievement:

- focus on academic subjects;
- graduation requirements;
- graduation rates;
- enrolment in rigorous programmes;
- recognition of academic accomplishments;
- academic expectations for students;
- uninterrupted class instruction;
- administrative involvement;

- quality and type of homework;
- teacher evaluation emphasizing learning.

Professional teaching conditions:

- teacher salaries;
- pupil load/class size;
- teacher time for planning;
- collegial work;
- teacher involvement in decision-making;
- teacher certainty;
- teacher autonomy/flexibility;
- administrative support for innovation;
- clerical support.

Although such a comprehensive list of context factors serves to emphasize the complexity of schools, it also has its limitations. Single items have little impact by themselves, and even when items are taken together they are better regarded as "enablers" than as causes of student learning. Another issue is that, as Oakes says, the "context features that are most easily recognised, measured, and reported may be the least likely to provide useful insights into school quality" (p. 195). Her list, she claims, is "elusive, complex and sometimes intangible" (p. 195), and she points to the difficulties of measurement when using the measures within an accountability system. The different nature of many of the factors in the list should also be noted. They seem to be an *ad hoc* collection, some of which are descriptive, others strategic, and not all of them open to change.

These problems apart, the most serious limitation of this collection of indicators is that it lacks an improvement or strategic dimension. Oakes' paper is more concerned with accountability than with improvement, and this may explain the nature of the list and its emphasis on measurement. The question should be: "How do these 'enablers' promote quality?" Unfortunately, this question is not asked often enough.

The review so far shows a great deal about the factors associated with school process but leaves many questions unanswered. What combinations of factors work best? What are the key factors? What is the primary focus of improvement? All of these questions are of central importance for understanding the process of improvement in schools. As Fullan maintains, "It is extremely important to define the ultimate focus of improvement in order to identify, classify and clarify the interrelationship of the main categories of variables." (1988, p. 26)

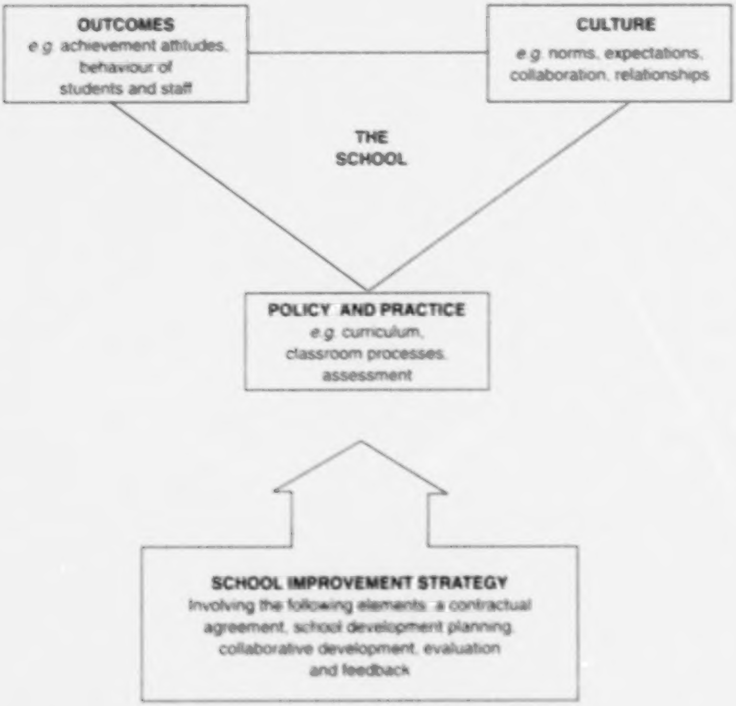
Fullan (1988) identifies four types of factors in school improvement. Background factors such as location, intake, or buildings are "givens". These are the features of schools that are unlikely to change in the short or medium term and therefore provide the context within which change occurs. Organisational variables relate to the internal features of the school's organisation, such as the level of staff collaboration, the style of leadership, and staff morale; they are more amenable to change but are also "given" at any particular point in time. Implementation strategy, in Fullan's terms, is a strategy used in a specific improvement project such as staff development. Outcomes are the result of the three previous factors.

The key to improvement is the implementation strategy, which must affect the background and organisational factors as well as the outcomes. It may need to be more or less powerful, depending on the relative "strength" of the other factors. In some cases, for example, the organisational variables may need to be changed or influenced by the implementation strategy. This is problematic because most implementation strategies have an instructional focus but must address organisational factors, which are often the main inhibitors of change. More powerful and integrative implementation strategies that directly address school culture are required. Fullan's analysis offers a dynamic alternative to the static conceptions of process previously reviewed, although even here linear, unidirectional relationships predominate.

It may be helpful to think of process in a more interactive way, as illustrated in Figure 8.1. As shown in the diagram, "process" is viewed as having four dimensions:

- the implementation strategy;
- the school's practices and policies;
- the school's organisational structure or management arrangements;
- the outcomes these produce not just for staff and students, but also for the culture of the school.

Figure 8.1 **A holistic approach to school improvement**


OUTCOMES
e.g. achievement attitudes, behaviour of students and staff
CULTURE
e.g. norms, expectations, collaboration, relationships
THE SCHOOL
POLICY AND PRACTICE
e.g. curriculum, classroom processes, assessment
SCHOOL IMPROVEMENT STRATEGY
Involving the following elements: a contractual agreement, school development planning, collaborative development, evaluation and feedback

BEST COPY AVAILABLE

Figure 8.1 proposes a "holistic" approach to process and school improvement that, although more complex, may also be more realistic. In the next part of this chapter, these dimensions are used as organising categories for a more detailed discussion of process indicators.

Summary

From one perspective, process indicators translate inputs into outputs; from another, they represent the constellation of factors that give a school its uniqueness. Whatever one's particular view, there is general agreement that "process" is the crucial determinant of quality within a school system. Process indicators are a proxy for those internal features of schools that "add value" to student ability. This definition restricts the discussion to those variables that are within or impinge directly on the school.

A review of research studies on school effectiveness and school improvement reveals a certain degree of confusion about the nature of process. By taking an improvement perspective, it is possible to divide "process" into four categories: policies and practices, school organisation or management arrangements, culture, and strategies. The review of the literature undertaken so far suggests a range of indicators under each of these headings.

Policies and practices:

- emphasis on curriculum and basic skills;
- monitoring student progress;
- clear goals;
- opportunity to learn;
- instructional time;
- structured teaching;
- classroom organisation;
- homework;
- communication and consensus.

Organisation:

- strong leadership;
- teacher involvement in decision-making;
- parental involvement.

Culture:

- high expectations;
- supportive and orderly climate;
- shared value system.

Strategies:

- planning;
- staff development;
- school review;
- external support.

There are obvious limitations to this list. No causal flow is implied and there is no indication of determining conditions. All indicators receive empirical support from the research literature, but there are conceptual difficulties in defining them operationally. All can be part of an improvement process as well as being the focus of an improvement effort. The next part of this chapter is concerned with the resolution of some of these difficulties.

Review of process indicators

Once again, important research studies relevant to process indicators are reviewed in order to refine the taxonomy presented in the summary above. Extensive citations from some recent studies are presented in order to give the reader an indication of the complexities involved.

Practices, policies, and organisation

By far the most extensive literature on school processes is the empirical research on teaching effects. Consistently high levels of correlation are achieved between student achievement scores and classroom processes (Brophy and Good, 1986; Walberg, 1990). This is a very complex territory, the intricacies of which are beyond the scope of this chapter, but one general conclusion stands out: "The most consistently replicated findings link achievement to the quantity and pacing of instruction" (Brophy and Good, 1986, p. 360); however formulated – "opportunity to learn", "instructional time", "time on task", "academic learning time" – this appears to be a crucial process variable.

Instructional time is, in itself, not a sufficient condition. The literature on teaching effects is replete with the tactics of effective instruction. Doyle (1987) provides a useful summary:

> Classroom studies of teaching effects have generally supported a direct and structured approach to instruction. That is, students usually achieve more when a teacher:
>
> - emphasizes academic goals, makes them explicit, and expects students to be able to master the curriculum;
> - carefully organises and sequences curriculum experiences;
> - clearly explains and illustrates what students are to learn;
> - frequently asks direct and specific questions to monitor students' progress and check their understanding;
> - provides students with ample opportunity to practice, gives prompts and feedback to ensure success and correct mistakes, and allows students to practice a skill until it is over-learned or automatic;
> - reviews regularly and holds students accountable for work.
>
> From this perspective, a teacher promotes student learning by being active in planning and organising instruction, explaining to students what they are to learn, arranging occasions for guided practice, monitoring progress, providing feedback, and otherwise helping students understand and accomplish work (p. 95).

But teaching is not just about short-term tactical responses. It is also about the consistent and strategic use of various teaching models (Joyce and Weil, 1986). A convincing body of research suggests that student achievement is enhanced by the use of particular teaching approaches. For example: "...the average student studying with the aid of organisers learns about as much as the 90th percentile student studying the same material without the assistance of the organising ideas" (Joyce *et al.*, 1987, p. 13), and, using the concept of effect size, it is also claimed that:

- School districts can now offer staff development programmes in the expectation that they will pay off in higher student achievement. Research conducted in the last ten years yields impressive evidence for the effectiveness of a variety of innovative teaching practices.
- Co-operative learning approaches, representing social models of teaching, yield effect sizes from modest to high. The more complex the outcomes – higher-order thinking, problem-solving, social skills and attitudes – the greater are the effects.
- Information-processing models, especially the use of advanced organisers and mnemonics, yield modest to substantial effect sizes; and the effects are long-lasting.
- Synectics and non-directive teaching, exemplifying personal models of teaching, attain their model-relevant purposes and influence student achievement in basic areas such as recall of information.
- DISTAR, an example of the behavioural family of models, yields modest effect sizes in achievement and, furthermore, influences aptitude to learn.
- When these models and strategies are combined, they have even greater potential for improving student learning (Joyce *et al.*, 1987, p. 13).

There is another aspect of teaching, which although of great importance to the lay person, is generally neglected by the research community; this is the impact of a teacher's personality on the individual student. In his attempt to construct a parsimonious list of performance indicators, Gray (1990) suggests that one of his three indicators of a "good" school would be the proportion of pupils who have a good or "vital" relationship with one or more teachers. Such a relationship can be regarded as a proxy for a very important aspect of the teaching process.

In terms of teaching, therefore, there seem to be four general process indicators worth considering: instructional time, teaching tactics, teaching strategies or models, and quality of teacher-student relationships. Taken together, they form a more comprehensive list of indicators than the previous one, and one that captures more fully the aspirations of school improvement.

The other major area of focus is the organisation of the school. "Strong leadership" is often cited as a key process variable, as are "teacher involvement in decision-making" and "parental involvement", but their operational implications are obscure and their empirical support unknown.

Parental and community involvement has to be a central process variable, in the light of common sense, contemporary values and ideas and current political views in most OECD countries. Further evidence of the importance of this variable emerges from Joyce's (1978) evaluation of the urban/rural project. The urban/rural programme was a six-year (1971-77) United States federal government experiment for improving education in poor communities. One of the outcomes of the shared professional/community governance of schools through decision-making councils was better learning by students. There is other evidence, but this study provides sufficient support for this relatively uncontentious, yet very important, process indicator.

"Leadership" and "teacher involvement in decision-making", even if they were more specifically defined, represent only the tip of the school's organisational iceberg. Some researchers now suggest that it is inappropriate to distinguish between classroom management and school management as if they were discrete areas of operation. The dialectic between them is a powerful indicator of school improvement. The term "empowerment" is increasingly being used to describe the main function of leadership in "healthy organisations". Gone are the images of strong leadership that eschews consultation, brooks no dissent and gives no thought to consensus or collaboration. "Leadership" is being redefined, the phrase "teacher decision-making" takes on a more contemporary meaning (Lieberman, 1988).

A helpful description of the style of leadership that links empowerment to school improvement is provided by Leithwood and Jantzi (1990) in their study of transactional leadership:

> "The study was prompted by evidence that variation in school cultures explains a significant proportion of the variation in staff practices and student outcomes across schools. Furthermore, one type of staff culture, which we have called 'collaborative', appeared to foster practices most conducive to the types of both student and staff development which are the focus of current school reform efforts. From previous work, we hunched that a conception of leadership as 'transformational' suggested strategies most likely to foster the development of collaborative cultures. After systematically assessing the extent of collaboration, we asked what strategies principals used to foster greater collaboration and were able to identify six. These included strengthening the school culture; using bureaucratic mechanisms; fostering staff development; frequent and direct communication; sharing power and responsibility and using rituals and symbols to express cultural values. The study provides support for the claim that principals have access to strategies which are 'transformational' in effect and, hence, assist in the development of collaborative school cultures. This means two things in our view: significant changes in staff members' individual and shared understandings of their current purposes and practices; and an enhanced capacity to solve future professional problems, individually and collegially." (p. 276)

Work on school development planning suggests that there are three dimensions to the 'management arrangements" in schools that are well adapted to school improvement (Hargreaves and Hopkins, 1991). The following descriptions of these aspects of management arrangements are drawn from schools where development planning is most successful. They are also consistent with the descriptions of the effective school described earlier.

The first consists of frameworks that guide the actions of all those involved in the school, without which the school runs the risk of lapsing into confusion and conflict; examples are the school's aims and policies and the systems for decision-making and consultation. The second concerns the clarification of roles and responsibilities, as all those involved in the school need to have a shared understanding of their respective roles and responsibilities; well-designed frameworks are useless without this understanding. The third is the promotion of ways in which the people involved can work together so that all find their role enjoyable and rewarding, as they work to achieve the aims of the school as a whole.

It may be helpful, therefore, to dispense with vague terms such as "strong leadership" and narrow conceptions such as "teacher decision-making" and replace them with an indication of the organisational functions to be acquired, such as establishing frameworks, clarifying roles and responsibilities, and working together. The key point is the necessity to see clearly what these functions are, not who performs them. Whilst the school leader undoubtedly plays a key role, management and the arrangements to support it are a collective activity and responsibility. This underscores the notion that there is a strong connection between management and culture.

The culture of the school

For the purposes of this chapter, the culture of the school is regarded as a kind of "meta process" indicator, that for reasons already given, is a vital yet neglected dimension of the improvement process. Although intangible, it holds the key to improving quality. It is a reflection of the norms and values of its members; it is the way they get things done. It is actively, though often unwittingly, constructed by the school's participants. The characteristics of schools as social institutions create a particular ethos – a set of values, attitudes and behaviours – which is representative of the school as a whole.

As De Caluwe *et al.* (1988) point out, the school's educational values are represented in tangible form in the school's organisational structures. Although the term "management arrangements" describes a particular organisation, it inevitably represents only a partial view. At a minimum, there must also be other factors already identified: high expectations, consensus on values, and an orderly and secure environment.

While these factors give an indication of the type of school culture that is supportive of school improvement, they do not provide a sufficiently comprehensive view. As Nias (1989) points out, despite the increasing number of studies to have examined the relationship between institutional cultures and the difficulty of educational change (Sarason, 1982; Rudduck, 1991), such studies "are unlikely to advance much further until the notion of culture is more carefully explored and given a stronger empirical base." (p. 143)

One of the main difficulties in translating research on school culture into indicators that identify processes conducive to school improvement is the frequent variation in the focus of the research. Thus, Hargreaves (1967) largely studied student cultures, Lortie (1975) examined cultures of teaching, and Deal (1985) described cultural symbols at the school level. There are many other excellent studies of school cultures, but they need to be more integrated and include multi-level studies that focus on the concerns of school improvement. This is particularly true at the interface between the classroom and the school; studies by Nias *et al.* (1989) and Heckman (1987), for example, have noted that some schools are renewing at the organisational but not at the classroom level.

It may thus be useful to look at two studies that have attempted multi-level analyses of school culture. Evans and Hopkins (1988) demonstrated a positive relationship between teacher personality, teacher innovativeness in the classroom and school culture. In the study, teachers who were operating at a low psychological level and in a "closed" school environment implemented few of the educational ideas and practices acquired on an exemplary and sustained in-service course. In contrast, teachers who were operating at a high psychological level and in an "open" school climate and had received the same training, implemented the ideas at a rate four times greater. In effect, the study suggested that the more open and democratic the school climate, and the more self-actualising the members of the teaching staff, the more effective their use of educational ideas in practice (Hopkins, 1990).

Similarly, Rosenholtz (1989) found that the social organisation of the school is seen to affect directly the commitment of teachers and the achievement of students. He contrasts the differences in teacher commitment in what are called "moving" and "stuck" schools (pp. 209-210) and demonstrates that the positive and negative consequences of workplace conditions are affected – if not determined – by the behaviour of the principal and district level policies and actions.

Although this discussion has demonstrated the importance of the culture of the school, it still remains difficult to define it in terms of operational process indicators. However, the factors "teacher self-esteem", "teacher commitment", and "teacher collaboration" are certainly relevant. It should also be reiterated that school culture is not just a "meta process indicator" but also a prime process indicator. Studies continue to demonstrate the impact of culture on student achievement and teacher behaviour (Joyce, 1990).

Strategies

Although variables such as those previously discussed are usually described individually, their impact on student and teacher outcomes is holistic rather than separate. They are also embedded in a school culture. How does one begin to affect these factors and the school culture in order to improve the quality of schooling? Unfortunately, as Fullan (1988) has demonstrated, most strategies for school improvement tend to address single factors or innovations rather than whole school issues. When strategies encompass whole school or cross-curricular issues, they are generally school-wide rather than school-deep. Fullan comments:

> "Without a direct and primary focus on changes in organisational factors it is unlikely that [single innovations or specific projects] will have much of a reform impact, and whatever impact there is will be short-lived... School improvement efforts which ignore these deeper organisational conditions are 'doomed to tinkering'... Strategies are needed that more directly address the culture of the organisation." (p. 29)

The point is well taken, but how does it apply to the strategies already identified? There is strong empirical support for "evolutionary planning" (Louis and Miles, 1990), staff development (Joyce and Showers, 1988), school review (Bollen and Hopkins, 1987), and external support (Louis and Loucks-Horsley, 1989) as school improvement strategies. There is also some indication that they are even more powerful in combination. An evaluation of projects that linked school review to teacher evaluation, for example, showed major improvements in a school's curriculum and instruction in a relatively short period of time (Bollington and Hopkins, 1989). School development planning, in particular, appears to be a strategy that can link together a variety of school improvement approaches, at least from the practitioners' viewpoint (Hargreaves and Hopkins, 1991).

When embedded in a school's organisation, these strategies combine to form an internal infrastructure that supports the management of change. But they are relatively crude indicators of organisational health. There are many schools that claim to have development plans or undertake reviews, and yet have improved little as a result. There

BEST COPY AVAILABLE

can be many reasons for this. The review may be regarded as a purely formal exercise, the culture may not change, or the internal dynamics may not be right. A number of recent studies shed light on this "micro-political" or interpersonal aspect of school improvement.

One is Rosenholtz's (1985) research on the organisational conditions of teaching:

> "Principals of effective schools have a unitary mission of improved student learning, and their actions convey certainty that these goals can be attained... Because the work of these principals pivots around improving student achievement, teachers have specific, concrete goals towards which to direct their efforts and know precisely when those efforts produce the desired effects. They are further encouraged by a supportive collegial group that lends ideas and assistance where needed. In turn, by achieving goals of student learning, teachers are provided with necessary motivation to continue to produce." (p. 352)

Another is from Corbett and Rossman's (1990) detailed qualitative analysis of the successful implementation of change:

> "First, certain antecedent conditions set the stage for how well or poorly a change project will go. Manipulating these to support innovative efforts creates an organisation capable of intentionally changing whenever a worthwhile opportunity presents itself.
>
> "Second, several intervening variables can be very powerful components of a change strategy. The three pivotal leverage points in the network seem to be: the encouragement/assistance, trial run and judgement of fit loop in the technical path; altering rules and procedures to accommodate change in the political path; and encouraging acceptance of new norms in the cultural path. The common denominators among the three were that at least some technical information was shared and systematic discussion of the information and trial runs of the new practices took place.
>
> "Essentially, then, implementation is greater in social, supportive settings than in isolated environments... Forcing teachers to implement directly as the result of a change in rules or procedures creates problems later." (pp. 187-188)

Finally, Fullan (1988), reviewing change in secondary schools suggests that the dynamics of change involve:

> "...a combination of factors which traditionally have been seen as separate or mutually exclusive. To cite three examples: active (aggressive) initiation followed by or coupled with progressive and widening collaboration and empowerment seem to be a powerful combination; both pressure and support are essential for success; the constellation of ownership, skill, mastery and commitment is more accurately portrayed as a phenomenon which builds throughout the change process rather than something which exists or is settled at the early stages." (p. 26)

These views suggest that single change strategies should not be too enthusiastically embraced and that careful attention should be paid to the dynamics of the processes involved. From the research reviewed so far, relevant elements appear to be:

- attention to cultural norms;
- clear goals with the focus on fundamental issues;
- explicit links to teaching;
- supportive and collaborative relationships with other teachers;

- pressure, often in the form of "active initiation", and high expectations from colleagues;
- frameworks or structures that promote practice, experimentation and empowerment;
- the motivation provided by competence, ownership and success.

Current rallying cries such as restructuring, development planning, site-based management, devolved financial management, and decentralised control are fine as far as they go. However, as Lieberman and Miller's (1990) article on restructuring illustrates, they will only result in school improvement if they embrace the dynamic aspects of the change process just described as well as broad strategies – such as self-review, action planning, staff development – that link together the classroom and the school. It is the power of these social-psychological processes to affect teacher behaviour that are the real predictors of enhanced outcomes for students.

The external perspective

The quest for school improvement is not confined to the school. Despite the tendency towards decentralisation in many OECD countries, school improvement is not a function of autonomy. The Dutch work that preceded the OECD International School Improvement Project (ISIP) referred to "the relatively autonomous school". More recent commentators have seen the school as the centre, rather than the unit of change. The distinction is important, because despite contemporary political pressure, research suggests that effective schools have a collaborative relation with outside agencies. Some recent work may help establish process indicators for the school's immediate external environment. The discussion is restricted to the district level, not because national policies do not affect school improvement but because their impact is uneven and not amenable to local control.

The school district interface has only recently become an object of study. Purkey and Smith (1985) prescribed four general policy recommendations for effectiveness:

- take the school as the focus of change;
- review the school situation;
- provide resources, technical assistance and training and encourage collaboration;
- establish appropriate shelter conditions at the local/district level.

This general recipe was corroborated by subsequent studies. Fullan's (1985) detailed and characteristically incisive review extends the work of Purkey and Smith, and the empirical study of school districts in Ontario (Fullan *et al.*, 1986) contributes a number of insights:

> "Neither grass-roots nor top-down approaches work by themselves. Central co-ordination, pressure, and development [are] essential, but so is corresponding school-based development on the part of principals and teachers as implementation decision-makers. The solution is neither more nor less centralisation, but rather it lies in the area of increased interaction and negotiation between schools and central offices, and investment in the development of capacities at both levels. The process is probably more powerful if it is initiated from the centre because the centre has more scope, resources and, consequently, potentially more influence. Once started,

equal attention must be given to development at both levels and to their co-ordination.

> "...once school-level development occurs, as it must if improvements are to be made, the school gradually takes on more initiative in not only identifying needed implementation resources, but also in selecting priorities. While the district may set up ways to enable and ensure that schools are focusing on implementation, we should not be mistaken about what is happening. It is a process of empowering the school and community and of committing district resources to follow through on implementation requests arising from school-level planning." (p. 325, 327)

Louis and Miles (1990) reach a similar conclusion:

> "District offices will have to learn to rely more on their working relationships with schools to steer a course through the turbulent waters, and less on rules and mandates. When there is pressure, it had better be accompanied by plenty of support. Schools have to have room, a good deal of local decision power, and help with the problems they face. That means a well-coupled relationship, not a distant one." (p. 291)

All three studies are describing a symbiotic relationship between school and district that is at the same time both "loose" and "tight" and generates a familiar constellation of factors:

- shelter conditions at each level;
- interaction and negotiation;
- imaginative resource allocation;
- capacity building and empowerment;
- pressure and support, including active initiation, sometimes from above.

"Loose-tight" coupling can take many forms. A brief look at how this relationship works in three different settings may make it possible to add some detail to the specification of factors.

The first example is a school/university/district project in the greater Toronto area. The initiative is based on an explicit model of classroom/school improvement that contains many of the process features already described in this chapter. The programme focused on classroom practice but recognised that practice cannot be sustained without support from both inside and outside the school. Some of the main partners in the project describe it as follows:

> "To increase the chances that the teachers would successfully transfer their new learning to the classroom, we built the programme to include certain elements. First, a powerful model of teaching was employed: co-operative learning. Second, we used an effective training strategy that provided follow-up support – the skill training model. Third, we combined co-operative learning training with instruction on implementing change. And fourth, volunteer participants were selected to participate on the basis of their interest in instructional improvement." (Fullan *et al.*, 1990, p. 18)

The second example comes from a project in Richmond County, Georgia, in which, by restructuring the "workplace conditions" for teachers, a school improvement programme obtained positive change in student achievement. The main collaborators describe the structure of the programme as follows:

> "We began our planning in January 1987 with intensive seminars for cabinet level staff. By March, district administrators had decided on the general dimensions of the project. During the first two years, the consultants would provide most of the training, but a cadre of teachers and administrators would be trained to offer service to other teachers and administrators – to bring other schools into the project on a regular basis in the future ...
>
> The development of the district cadre was critical to the project and to the relationship between the district and the consultants; it symbolised the intent to make permanent changes in the workplace. It made concrete the need for district personnel to possess the expertise of the consultants and to take over the functions of the consultants." (Joyce *et al.*, 1989, pp. 71-72)

The third example describes an attempt to improve schools on the basis of school evaluations conducted by local inspectors in London. In English schools, teams of inspectors traditionally visit a school for a week and then produce reports which are often published. However, this information is rarely used systematically for school improvement purposes. A recent Chief Inspector in the Inner London Education Authority (ILEA) regarded this situation as unsatisfactory, and despite thorough inspections, he judged that:

> "There remained the important task of how these schools might be improved to narrow the gap between the best and the worst. It was evident that such schools needed sustained help and support, yet inspectors had insufficient time at their disposal to work intensively with them. The solution was a team of inspectors who would be freed from all normal duties to devote all their time to school improvement. This became known as the IBIS or Inspectors Based in Schools scheme. In the secondary phase a team of twelve inspectors undertook this task, beginning in autumn 1986. Entry to the school was for the most part amicably negotiated between the Inspectorate and the Principal. The team introduced themselves at a staff meeting, explained the team's approach and answered questions. The life of the school and lessons were observed by the team, who also interviewed all the staff as well as some pupils and parents. The team then withdrew for a few days to write a diagnostic report, which was presented to the staff as a series of discussion documents. Following a full discussion with the staff, the team remained in the school for a further four weeks or so to engage in the developmental work arising out of the diagnostic reports and the discussion of them. At the end of this phase, the team wrote its final report for the school's governors and this covered both the school's strengths and weaknesses (which is somewhat like an inspection report) and the action being taken in the developmental phase to remedy the weaknesses (normally not part of an inspection report)." (Hargreaves, 1990, pp. 18-19)

These initiatives illustrate a number of different facets of successful school/district collaboration and "loose coupling" in action, principally the need for:

- powerful and ongoing classroom-based staff development;
- an explicit focus on teaching, informed by research;
- decisions and judgements made on the basis of evidence;
- establishment of a cadre to provide ongoing support, empowerment and capacity building;
- a medium or long-term view of the change process, coupled to a coherent strategy;
- contractual obligations for all participants.

Summary

On the basis of what precedes, a list of process indicators can be developed that provides some indication of the characteristics of an external environment supportive of school improvement, and the list of process indicators presented at the end of the first part of this chapter can now be refined.

Strategies:

- planning;
- staff development;
- school review;
- external support, preferably in collaboration with staff, and with particular attention to the dynamics of the process, *e.g.*:
  - clear links to teaching;
  - commitment to practice and experimentation;
  - intrinsic motivation;
  - pressure and support.

Policies and practices:

- instructional time;
- teaching tactics;
- teaching models;
- quality of student-teacher relationships.

Organisation:

- establishing frameworks;
- clarifying roles and responsibilities;
- ways of working together;
- involving parents and community in the school.

Culture:

- high expectations;
- supportive and orderly climate;
- shared value system;
- teacher self-esteem;
- teacher commitment;
- teacher collaboration.

External relationships:

- ongoing staff development;
- contractual obligations;
- establishing a cadre for capacity building and ongoing support;
- medium/long-term process of change;
- shelter conditions at each level.

Applications of process indicators

International perspectives

It is difficult to address process indicators from an international perspective, because the research literature is largely Anglo-Saxon and often quantitative in orientation. However, the extensive literature originating from the ISIP project is generally supportive of the types of indicators discussed here. All of the macro strategies described in the previous section, for example, have been used successfully in a number of different OECD countries (van Velzen *et al.*, 1985; Hopkins, 1987).

Different interpretations are given to various concepts in different countries; leadership is a prime example. It has been said that cross-cultural comparisons of leadership, between, say, England and Denmark, are impossible, because the former has a long tradition of autonomous headteachers, whereas the role of school leader is unrecognisable in the latter. This may be true, but only if one focuses on role. If the concern is for function, then leadership is present in all successful schools, especially in view of the increasing interaction between classroom and school management. Admittedly, this function may be difficult to measure, but it is amenable to observation.

The INES project has already undertaken some excellent work in this area. The four indicators identified by Network C (Ballion, 1990), *i.e.* leadership, staff collegiality, curriculum and success-oriented policy, have already been subjected to comprehensive international comparison by members of this working group.

Another way of gauging the cross-cultural transferability of indicators is to see how they apply to the education of minority ethnic groups within a dominant cultural setting. Effective schools research is a case in point. Cummins (1986), for example, proposes a theoretical framework of four elements for improving the academic achievement of minority students:

- incorporating language and culture into school programmes;
- involving minority parents as partners;
- characterising effective instruction as reciprocal interaction;
- implementing advocacy-oriented assessment.

Stedman offers an alternative formula (1987) for effective schools for ethnic minority pupils:

- cultural pluralism;
- parent participation;
- shared governance;
- academically rich programmes;
- skilled use and training of teachers;
- personal attention to students;
- student responsibility for student affairs.

There is a high degree of consistency between the characteristics of the effective school and the specifications of school cultures that are advanced as being supportive of enhanced outcomes for students from ethnic minority backgrounds. This lends support to the position that in Western countries, particularly in terms of function, process indicators transcend cultural boundaries, but it does not resolve the difficulty of measurement. It may not be possible to construct a reliable survey instrument on the basis of these indicators, but it may be possible to use observation and the exercise of professional judgement to compare and thus to measure these factors.

Using process indicators

Despite the empirical support, this collection of process indicators for school improvement is at best a map of the territory. In no sense can it claim to provide causal links to student achievement. In terms of school improvement, the theory only becomes useful when it is put into practice. As Stenhouse (1975) noted in a slightly different context, such proposals are not to be regarded "as an unqualified recommendation but rather as a provisional specification claiming no more than to be worth putting to the test of practice. Such proposals claim to be intelligent rather than correct" (p. 142). This, of course, is the purpose of any taxonomy of process indicators.

Such a collection of indicators makes it possible to appreciate more fully the complexities of the schooling process. There is of course no intention that this or any other similar list should be used as the basis of a quantitative survey instrument. Its use, as Scheerens (1990) argues, is to complement performance indicators. But the judgements made on the basis of such criteria can be aggregated, quantified, and used as a basis for comparison. It is necessary, of course, that the criteria be adequately conceptualised and negotiated, and that their meaning be clear. The evidence on which judgements are based also needs to be adduced, but this is no more than good practice. Seen in this way, these indicators provide a necessary component of any improvement process.

The list of process indicators developed in this chapter can be used for self-evaluation and diagnosis at the classroom and school levels. Discussion should always precede major planning efforts, and process goals need to be part of any reform programme. At the local level, the indicators can be used as the basis for evaluation, and/or inspection, as in the IBIS example. They can also help to determine the process elements necessary for any local school improvement initiative. At the national level, they can be used for monitoring, and as a means of informing policy initiatives. They can perhaps provide the basis for international discussion, if not comparison, particularly when the focus is on function rather than role.

One of the clear messages from school improvement research is the importance of a guiding vision. Above all, the indicators can help schools to acquire a vision of process that can lead to improved outcomes for all its pupils.

Conclusion

It is now over ten years since Edmonds (1978) asked, "How many effective schools would you have to see to be persuaded of the educability of all children?" He continued, "We already know more than we need to do that. Whether or not we do it must depend on how we feel about the fact that we haven't so far." Despite increasingly sophisticated knowledge about the process of educational change, the school as an organisation, and the various strategies for improvement (Fullan, 1991), student achievement still lags far behind society's expectations. The reasons for this are legion, but three themes stand out.

First, a major difficulty seems to be the way in which school improvement knowledge is used. Knowledge of the type discussed in this chapter is at best informed advice that schools may wish to test in their own situations, and policy-makers may wish to incorporate into their mandates. The advantage of school improvement strategies is that they provide a means for putting this knowledge into practice. The knowledge is not there to control, but to inform and discipline practice. This is the way in which process indicators of school improvement should be used.

Second, although a lot is known about school improvement, there remains an element of serendipity in achieving educational quality. Schools are highly resistant to external pressure to change. Joyce (1990) captured this paradox nicely when he said that "educational change is technically simple but socially complex". It is the social complexity that militates against neat categorisation and prescription. Yet this is why continued attention must be given, both in policies and practices, to the social organisation of schools and to how they create their own cultures.

Finally, a word about politics. Educational change is usually the result of a political process at both the macro and micro levels. On the macro level, centralised policies create the agenda, whereas at the micro level implementation determines outcomes. The implementation of policy is unpredictable, because neither of these levels has, in reality, much influence on the other. Schools cannot determine national or local policy, although they can decide what they want to do. Similarly, although policy-makers may set the agenda, they cannot control outcomes, because the process of implementation is mediated through many different school contexts. Unfortunately, most systems fail to realise that educational quality is a function of the dialectic between policy and practice, not the preserve of one or the other, although both have their part to play and their role to fulfil. It is time to recentre the debate on the process of schooling. It is apparent from the research reviewed in this chapter that student achievement is positively related to school processes. It is here that policy is translated into practice and that the work of policy-makers and practitioners come together. It is with the specification of process indicators for school improvement that the dialogue should begin.

Acknowledgements

I am most grateful to Mel Ainscow and July Howard for helpful comments on an earlier draft and to July Howard and Angela Cutts for assisting with the ERIC search undertaken for the second part of this chapter.

Bibliography

BALLION, R. (1990), "Analytical Dimensions in School Functioning", paper prepared for OECD/CERI INES Project meeting, November (mimeo).

BOLLEN, R. and HOPKINS, D. (1987), *School Based Review: Towards a Praxis*, ACCO, Leuven, Belgium.

BOLLINGTON, R. and HOPKINS, D. (1989), "School Based Review as a Strategy for the Implementation of Teacher Appraisal and School Improvement", *Educational Change and Development*, Vol. 10 (4), pp. 8-17.

BROPHY, J. and GOOD, T. (1986), "Teacher Behavior and Student Achievement", in Wittrock, M. (ed.), *Handbook of Research on Teaching* (3rd edition), Macmillan, New York, pp. 328-375.

CORBETT, H.D. and ROSSMAN, G. (1989), "Three Paths to Implementing Change", *Curriculum Inquiry*, Vol. 19 (2), pp. 163-190.

CUBAN, L. (1983), "Effective Schools: A Friendly but Cautionary Note", *Phi Delta Kappan*, Vol. 64 (10), pp. 695-696.

CUMMINS, J. (1986), "Empowering Minority Students: A Framework for Intervention", *Harvard Educational Review*, Vol. 56 (1), pp. 18-36.

DE CALUWE, L., MARX, E. and PETRI, M. (1988), *School Development: Models and Change*, ACCO, Leuven, Belgium.

DEAL, T. (1985), "The Symbolism of Effective Schools", *The Elementary School Journal*, Vol. 85 (3), pp. 601-620.

DOYLE, W. (1987), "Research on Teaching Effects as a Resource for Improving Instruction", in Wideen, M. and Andrews, I. (eds.), *Staff Development for School Improvement*, Falmer Press, Lewes, Sussex, pp. 91-102.

EDMONDS, R. (1978), "A Discussion of the Literature and Issues Related to Effective Schooling", paper prepared for the National Conference on Urban Education, Cemrel, St. Louis, June.

EVANS, M. and HOPKINS, D. (1988), "School Climate and the Psychological State of the Individual Teacher as Factors Affecting the Use of Educational Ideas Following an Inservice Course", *British Educational Research Journal*, Vol. 14 (3), pp. 211-230.

FULLAN, M. (1985), "Change Processes and Strategies at the Local Level", *The Elementary School Journal*, Vol. 85 (3), pp. 391-421.

FULLAN, M. (1988), "*Change Processes in Secondary Schools: Towards a More Fundamental Agenda*", University of Toronto, Ontario (mimeo).

FULLAN, M. (1991), *The New Meaning of Educational Change*, Cassell, London.

FULLAN, M., ANDERSON, S. and NEWTON, E. (1986), *Support Systems For Implementing Curriculum In School Boards*, Ministry of Education, Toronto, Ontario.

FULLAN, M., BENNETT, B. and ROLHEISER-BENNETT, C. (1990), "Linking Classroom and School Improvement", *Educational Leadership*, Vol. 47 (8), May, pp. 13-19.

GRAY, J. (1990), "The Quality of Schooling: Frameworks for Judgement", Inaugural Lecture, QQSE Research Group, University of Sheffield, March (mimeo).

HARGREAVES, A. (1991), "Cultures of Teaching: A Focus for Change", in Hargreaves, A. and Fullan, M. (eds.), *Understanding Teacher Development*, Cassell, London.

HARGREAVES, D.H. (1967), *Social Relations in a Secondary School*, Routledge and Kegan Paul, London.

HARGREAVES, D.H. (1990), "Accountability and School Improvement in the Work of LEA Inspectorates: The Rhetoric and Beyond", Department of Education, University of Cambridge (mimeo).

HARGREAVES, D.H. and HOPKINS, D. (1991), *The Empowered School*, Cassell, London.

HECKMAN, P. (1987), "Understanding School Culture", in Goodlad, J. (ed.), *The Ecology of School Renewal*, 86th Yearbook for the National Society for the Study of Education, NSSE, Chicago, IL.

HOPKINS, D. (1987), *Improving the Quality of Schooling*, Falmer Press, Lewes, Sussex.

HOPKINS, D. (1990), "Integrating Staff Development and School Improvement", in Joyce, B. (ed.), *Changing School Culture Through Staff Development*, 1990 Yearbook of the Association for Supervision and Curriculum Development, Alexandria, VA, pp. 41-67.

JOYCE, B. (1978), *Involvement: A Study of Shared Governance in Teacher Education*, National Dissemination Center, University of Syracuse, Syracuse, NY.

JOYCE, B. (ed.) (1990), *Changing School Culture Through Staff Development*, 1990 Yearbook of the Association for Supervision and Curriculum Development, Alexandria, VA.

JOYCE, B., SHOWERS, B. and ROLHEISER-BENNETT, C. (1987), "Staff Development and Student Learning: A Synthesis of Research on Models Of Teaching", *Educational Leadership*, Vol. 45 (2), pp. 11-23.

JOYCE, B. and SHOWERS, B. (1988), *Student Achievement Through Staff Development*, Longman, New York.

JOYCE, B., MURPHY, C., SHOWERS, B. and MURPHY, J. (1989) "School Renewal as Cultural Change", *Educational Leadership*, Vol. 47 (3), pp. 70-77.

JOYCE, B. and WEIL, M. (1986), *Models of Teaching* (3rd edition), Prentice-Hall, Englewood Cliffs, NJ.

LIEBERMAN, A. (1988), "Teachers and Principals: Turf, Tension, and New Tasks", *Phi Delta Kappan*, May, pp. 648-653.

LIEBERMAN, A. and MILLER, L. (1990), "Restructuring Schools: What Matters and Why", *Phi Delta Kappan*, June, pp. 759-764.

LEITHWOOD, K. and JANTZI, D. (1990), "Transformational Leadership: How Principals can help Reform School Cultures", *School Effectiveness and School Improvement*, Vol. 1 (4), pp. 249-280.

LORTIE, D. (1975), *Schoolteacher*, University of Chicago Press, Chicago.

LOUIS, K.S. and LOUCKS-HORSLEY, S. (1989), *Supporting School Improvement: A Comparative Perspective*, ACCO, Leuven, Belgium.

LOUIS, K.S. and MILES, M. (1990), *Improving the Urban High School*, Teachers College Press, New York.

MORTIMORE, P., SAMMONS, P., STOLL, L., LEWIS, D. and ECOB, R. (1988), *School Matters*, Open Books, London.

NIAS, J. (1989), "Refining the Cultural Perspective", *Cambridge Journal of Education*, Vol. 19 (2), pp. 143-146.

NIAS, J., SOUTHWORTH, G. and YEOMANS, R. (1989), *Staff Relationships in the Primary School*, Cassell, London.

NUTTALL, D., GOLDSTEIN, H., PROSSER, R. and RASBASH, J. (1989), "Differential School Effectiveness", *International Journal of Educational Research*, Vol. 13 (7), pp. 769-776.

OAKES, J. (1989), "Why Educational Indicators? The Case for Assessing School Context", *Educational Evaluation and Policy Analysis*, Vol. 11 (2), pp. 181-199.

PURKEY, S. and SMITH, M. (1983), "Effective Schools: A Review", *The Elementary School Journal*, Vol. 83 (4), pp. 427-452.

PURKEY, S. and SMITH, M. (1985), "School Reform: The District Policy Implications of the Effective Schools Literature", *The Elementary School Journal*, Vol. 85 (3), pp. 352-389.

ROSENHOLTZ, S. (1985), "Effective Schools: Interpreting the Evidence", *American Journal of Education*, May, pp. 352-388.

ROSENHOLTZ, S. (1989), *Teachers' Workplace*, Longman, New York.

RUDDUCK, J. (1991), *Innovation and Change*, Open University Press, Milton Keynes.

RUTTER, M., MAUGHAN, B., MORTIMORE, P., OUSTON, J. and SMITH, A. (1979), *Fifteen Thousand Hours*, Harvard University Press, Cambridge, MA.

SARASON, S. (1982), *The Culture of the School and the Problem of Change* (2nd edition), Allyn and Bacon, Boston.

SCHEERENS, J. (1990), "School Effectiveness Research and the Development of Process Indicators of School Functioning", *School Effectiveness and School Improvement*, Vol. 1 (1), pp. 61-80.

SCHEERENS, J. (1992), "Process Indicators of School Functioning", in OECD, *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, pp. 53-76.

STEDMAN, L. (1987), "It's Time We Changed the Effective Schools Formula", *PhiDelta Kappan*, November, pp. 215-224.

STENHOUSE, L. (1975), *An Introduction to Curriculum Research and Development*, Heinemann Educational Books, London.

TAYLOR, B. (1990), *Case Studies in Effective Schools Research*, Kendall/Hunt Publishing Company for the Center for Effective Schools, Dubuque, IO.

WALBERG, H.J. (1990), "Productive Teaching and Instruction: Assessing the Knowledge Base", *Phi Delta Kappan*, Vol. 71 (6), pp. 470-478.

van VELZEN, W., MILES, M., EKHOLM, M., HAMEYER, U. and ROBIN, D. (1985), *Making School Improvement Work*, ACCO, Leuven, Belgium.

BEST COPY AVAILABLE

Part Three

INDICATORS OF OUTCOMES OF EDUCATION

Chapter 9

Evaluating School Performance in France

by

Martine Le Guen
Ministry of Education, France

There has been a major effort since the 1960s to develop education, and the information needs of governments have grown correspondingly. Earlier demand for information on resource requirements has been complemented by a demand for the assessment of school and student performance.

Evaluation is today a central concern of not only education authorities but also of teachers, parents and researchers. Now that it benefits from sound methods and can provide valid and reliable results, it increasingly enjoys credibility in the community.

As an aid to understanding the French education system, the section below offers a brief recapitulation of the French primary and secondary education structure. A comparison of this structure with the International Classification of Education is given in the Annex.

Why student evaluation has become more important

Coping with the rapid expansion of public education was the main concern during the 1960s for governments trying to solve difficult logistic problems. To quote Jacques Lesourne:

> "It is striking, when reading international reports, to note the similarity in trends and concerns, despite the differences in the various countries' education systems. The expansion in the 1960s and 1970s shared common features everywhere: increased public spending on education, longer compulsory schooling, the spread of comprehensive secondary schools (symbol of democracy)..." (Lesourne, 1987)

During this period, managers experienced a crucial need for readily available quantitative information. To cope adequately with baby boom school entrants and the raising of the school-leaving age to 16, planners could not do without statistics on school-age populations, teacher requirements, wagebills, and the number of schools to build and operate. At the beginning of each school year, figure- and fact-filled dossiers were needed as a basis for hard-bargaining budget discussions. It is easy to understand why, between

1960 and 1979, when 72 per cent of French State secondary schools were built, accommodation had to take precedence over considerations of quality. It quickly became necessary, however, to obtain more refined knowledge of how pupils passed through the school system. A clearer picture was needed of both the length of schooling undergone by successive generations of pupils and the flows following the various channels of education.

In 1978/79, the Ministry for Education, Youth and Sport set up the first "pupil panel", in which a cohort of 20 000 children was systematically monitored over five years beginning with their enrolment in their first year (*cours préparatoire, CP*) of primary school. A similar panel was instituted in 1980, but beginning with enrolment in the sixth year, *i.e.* the first year of secondary school. The first finding was how important pre-schooling is to school performance, especially for pupils from underprivileged backgrounds or of foreign nationality. It was also found that, while the parents' socio-occupational profile played a large role in a child's chance of success, the crucial factor in success or failure remained the parents' own standard of education. The tracer studies revealed that a child's educational future is largely determined by success in primary school, the first year of Preparatory Class (CP) in particular. As Duthoit (1988) has pointed out, repeating the CP is a handicap which is difficult to overcome: "Repeating in CP leaves a pupil with only a four in ten chance of reaching the sixth grade without further failure, another repeat or falling by the wayside."

Yet this information, useful though it is, does not tell us about the real standard of pupils at key points during their school career. The only indications it provides concern a pupil's orientation, that is, the streaming decision taken, and secondary leaving certificate results. It was thus important to supplement the system with some form of "quality control" to give a more accurate idea of pupil performance.

The democratisation of education and the raising of the school-leaving age to 16 caused a surge in secondary school attendance. In France, the "Haby Reform" in 1977/78 introduced the *collège unique* (comprehensive high school dispensing a single programme of education), henceforth attended by all but a small minority of pupils. From the mid-1970s on, as a result, 95 per cent of each age cohort reached sixth grade compared with 43 per cent in 1960.

The old certificate system signposting the various stages of education was simultaneously phased out: the primary school certificate, the sixth grade entry examination, the lower secondary school (ISCED 2) certificate – from 1978 was awarded simply by in-school assessment, before being reintroduced with a mixture of assessment and national examinations. The effect of this was to heighten the role of the first compulsory examination – the upper secondary school (ISCED 3) leaving examination (*baccalauréat*), and to remove the intermediate benchmarks for measuring pupil proficiency.

Higher school participation thus gave rise naturally to concern about standards, and attention rapidly shifted from throwing open the schools to the quality of the education they provided. During the 1980s, a major effort was made to design a system for evaluating educational efficiency, ascertaining what pupils were actually learning. "The need for a forecasting and evaluation system is obvious, for three reasons. First, the size and complexity of schools rule out 'seat-of-the-pants'' navigation. Second, schools must be able to adjust to a changing scientific, technical, economic and social milieu. Third, they must be in a position to correct their own faulty organisation and, in particular, high failure rates" (Prost, 1983).

Moreover, as governments have continued to spend heavily on education, measurement of knowledge acquisition among pupils has become essential, particularly in the broader public debate. With the media feeding the public debate over "declining standards in education", taxpayers and employers are entitled to a clear view of how the school system performs. Parents, for their part, need accurate information about the strengths and weaknesses of their children in comparison with national standards.

Reinforcing these trends, the demand for education continues to grow. In 1982, 49 per cent of 18 year-olds attended school; by 1989, the figure had risen to 65 per cent. Families, and teenagers themselves, want schooling which will provide the right kind of diploma for obtaining a good job. Awareness of the importance for France of developing its education system is apparent in the broad themes of the Education Orientation Act of 10 July 1989: "Education is the top priority of the nation. France's objective over the next ten years is to make sure that all children in a given age group attain at least a vocational diploma, and that 80 per cent reach the *baccalauréat.*"

In addition, the move away from tight central control to greater local and regional autonomy has made regular evaluation of school performance more necessary to enable the authorities to gauge the achievement of policy goals. The education system must, if it is to measure progress in school performance, continually evaluate itself and use the findings to adjust its action. Nowadays, as Boisivon (1990) observes, "Education authorities and those concerned by the proper functioning of the system must insist on evaluation procedures being introduced; they are factors of progress and form part of the information system."

How does the community react to the need for evaluation? Until about ten years ago, evaluation was equated with teachers' reports. Now that anonymity is felt to be guaranteed and arbitrary decisions eliminated, most people today see evaluation as a tool for judging how well the education system is working. Evaluation is considered *legitimate*, as it provides valuable information. A teacher can situate the level of a class, or a headmaster can evaluate a school's standards in comparison with the national analysis (which is the only one published). Parents appreciate the effort being made to supply school performance data and make them more understandable. Whereas, a few years ago, the only means of judging the development of young people's learning was through the comparative assessment of military service induction tests, national evaluation techniques today offer the possibility of scientifically comparing results several years apart (Baudelot and Establet, 1988).

The principle of evaluation having been accepted, how has it been organised in France?

The nature of school evaluation in France

National assessment of pupil knowledge acquisition is carried out by a unit in the Central Administration of the French Ministry for Education, Youth and Sport. Set up in 1974, it was the brain-child of Joseph Fontanet, a French Minister of Education who had formerly been a company director. The circular No. 74-204 of 24 May 1974, which provided for the establishment of the assessment unit, states: "The Education Performance Evaluation Department is responsible for collecting, analysing and making available to the Minister the data required in order to be able to evaluate all aspects of the results of education and training and to monitor their trends."

The initial idea, which has endured, was to entrust the job of evaluation to a neutral unit, that is, one that had within the Ministry neither a control function nor teaching responsibilities. Outside assessors would not be used; rather the institution itself would be enabled to undertake the investigation without fear or favour. The stated aim was to make a dispassionate judgement on the education system's performance and particularly on what pupils actually learned.

The technical service – which has since become a Directorate of the Ministry – was therefore approached with a request to expand its existing quantity assessment system. The Directorate for Evaluation and Prospective Analysis, which now helps in the national evaluation of pupil knowledge acquisition, developed its methods in close co-operation and partnership with the Inspectorate-General of Education, the Inspectorate-General of Education Administration, teachers and researchers. Its survey findings provide information for public debate and social dialogue, and are used as a "navigational aid" by decision-makers.

National assessment of student learning

Established in February 1987, the Directorate for Evaluation and Prospective Analysis is a conglomerate of various technical services in the Ministry for Education, Youth and Sport whose main job it was to design, run and expand the Ministry's information system. The Directorate is not empowered to take educational policy or budget decisions. Its neutral status confers a degree of objectivity on the conduct of its surveys. Its technical facilities provide survey officials with direct access to statistical data bases, a real advantage when drawing samples or doing calculations requiring the use of file statistics.

As part of the Central Administration, the Directorate can devise circulars to be sent out to regional bodies and schools. While this cannot of course in itself ensure the success of evaluation exercises, which do not in any case constitute an obligation, it certainly does help, especially when combined with a persuasive information campaign.

For the sake of convenience and clarity, the Directorate for Evaluation and Prospective Analysis (DEPA) draws up an annual work programme enumerating the pupil knowledge surveys. Some of these are initiated by the Directorate; others are commissioned. The work programme is submitted for discussion to the Minister's Private Office, the Inspectorates-General and the other Directorates in the Ministry. After negotiation, it is finalised in accordance with the institution's particular concerns.

So far, the tasks of evaluation recently assigned to the Inspectorates-General has not given rise to overlap, since these are complementary: the DEPA records data-revealing trends that are measured as indicators; its surveys highlight the operational problems encountered. Using them as a basis, the Inspectorates-General carry out on-the-spot investigations to discover the causes of the facts observed.

Participation is necessary for successful evaluation

From the outset, it seemed advisable to associate the various interest groups with the survey operations. As two former evaluation officials acknowledge, this approach, while seemingly "slow-paced and sometimes expensive, appears correct and self-evident once it is realised that the findings must be accepted and endorsed by all the partners and, especially, used in teaching practice" (Mondon and Seibel, 1987).

Once a survey is decided upon, the next step is to create a national steering committee made up of representatives of the teaching Directorates concerned, the Inspectorate-General of Education, regional and departmental inspectorates, various teacher groups (according to the target level of the survey: primary school teacher, secondary school teacher, certified teacher, *agrégé*, etc.), researchers (often subject-area experts, or psychometricians) and DEPA specialists. The steering committee reaches agreement on the survey objectives and content. The diversity of the steering committee, composed as it is of people from a number of backgrounds, ensures its neutrality in regard to teaching theories. A survey is not an opportunity to "sell" a preferred teaching philosophy; its purpose is to observe the workings of the education system.

Participation is also a feature of the communication phase in which the representatives of the teachers' union and parent-teacher association are informed. They are notified of each new survey so that they can see for themselves that it is devoid of value judgements on individual teaching methods and unconnected with the streaming of particular pupils. These consultations are seen as beneficial, as the partners generally agree on the survey goals, even though the teachers complain about the extra workload involved.

A few days before the survey is launched, headteachers of the schools in the sample and teachers are briefed on the survey methods and purpose. The briefings provide an opportunity for dialogue with the partners in the field, who can comment on goals, methods and materials and make suggestions for improving the survey arrangements.

Sound methods ensure reliable findings

The essence of the method – which can be refined according to the problem treated – consists in determining objectives, testing questionnaires, providing safeguards and establishing representative samples of the national situation. These elements are briefly discussed below.

i) Objectives

The first job of the national steering committee is to define the objectives. Usually this means carefully studying official programmes and directives in order to identify the implied or explicit education activities requiring evaluation. Members of the steering committee often consider this to be a boring exercise, as they would prefer to get on with drafting the survey papers. But it is a crucial step, and must precede any evaluation since findings must be matched with a frame of reference if they are to be usefully interpreted. The list of objectives sets the limits of the investigation, for each objective identified requires a survey chapter and so determines the framework for analysing the results that shed light on the system's performance.

The work done to identify and operationalise the objectives has definite educational value. The clarification of programme goals and the spelling out of aims enables teachers to work more readily towards achieving those aims rather than being constantly bogged down in the everyday responsibilities of teaching.

There are occasions when the survey goes beyond what is taught in school, but here again the objectives must be defined. For example, when the Education and Economics Commission asked for a survey on the acquisition, at the end of lower secondary schooling, of the economic notions essential to a future citizen, it was impossible to use a teaching syllabus as a reference, since basic economics is not really taught before the first year of upper secondary school. The overall aim of this exercise was to discover how well ninth grade pupils perceived and understood the economic scene on the basis of general knowledge gleaned mainly from history and geography lessons, in the home or through the media.

Another recent survey operation aimed at gauging the ability of secondary school-leavers to understand and express themselves in English. Once more, the aim was not to make an assessment related to the curricula of the particular streams – which in any case do not have comparable class timetables – but rather to evaluate the degree of proficiency in a modern language of a pupil leaving secondary school after *n* years. University and professional specialists were consulted beforehand to help in determining "par" proficiencies.

ii) Piloting of survey instruments

Once the table of objectives has been completed, a working party begins drafting the survey papers. Each exercise is tried out on pilot classes. Its reliability is checked during this trial phase by testing the pupils' comprehension of the instructions, the time needed for completion, and the appropriateness of the questions. The trial also serves to standardize the survey instructions to be issued by the teachers of the sample classes and the marking procedure applied by teachers during the final evaluation according to a very precise handbook. Collection of data under standard conditions means that they can be statistically processed and condensed; conclusions can then be drawn concerning populations rather than individuals.

Trial-runs are also a way of making sure that the most effective exercises are selected. Factor analysis of cross-references and orders of importance avoid redundant questions and strengthen the structure of the survey. The comments of the teachers who take part in this phase provide a valuable contribution to improving the survey arrangements.

iii) Evaluation safeguards

All education surveys are protected by statistical secrecy under the terms of Act No. 78-17 of 6 January 1978. The processing of name-carrying data, even those which become anonymous in the final stage, must be reported to the National Commission on Computers and Privacy, which is charged with protecting civil liberties.

In the case of an education survey, the DEPA alone has access to data. Precautions are taken to avoid performance of a pupil, class, school or region being divulged. The national picture is provided by the evaluated schools taken as a whole, not by any single one of them. The national findings are a helpful yardstick which a teacher or principal can use to compare the pupils' or school's performance. It is worth repeating that the DEPA, because of the highly technical nature of its assignments, is an impartial evaluator. More specifically, evaluation avoids glorifying any particular theory of education; neither is it a pretext for checking a pupil's or a teacher's classroom performance. It is carried out after a lesson has been taught. The findings shed light on the overall efficiency of the education system.

iv) Sampling

A very large proportion of national evaluation work relies on poll survey techniques. Random samples of schools are drawn using two stratification variables: rural-urban environment as expressed by the size of the commune; and size of the school. In the schools selected, all pupils of the level being tested take part in the survey, so that principals will not be tempted to choose particular classes or individuals.

This method is generally preferred to scattered pupil sampling, since surveys usually try to capture the "establishment" dimension, with its peculiar strategies and practices, by means of a questionnaire set for principals and teachers.

v) Teacher involvement

Choosing to observe the reality of classroom education implies choosing to work in co-operation with teachers. The evaluation process (in its setting and marking stages) requires the participation of teaching *professionals*. The teachers must therefore take on the tasks of an evaluation in accordance with the instructions provided.

Teachers are required to carry out the instructions to the letter, even though these may not correspond at all to their usual teaching methods. They are not being asked to teach; they are being asked to collaborate in a nationally representative survey. There are times in an evaluation, of modern languages, for example, when teachers are asked to vary normal practice and to give pupils instructions in French, this being the only way to ascertain knowledge acquisition without distortion induced by problems of understanding.

Disaggregation

National evaluation of student learning relies on highly centralised procedures. The national steering committee and the task forces are organised by the DEPA. The tables of objectives and the survey instruments are worked out on the national level, even though in some recent instances (see below) local partners have been asked to contribute more. With the number of surveys increasing, and the intention of certain local authorities to conduct their own studies, it is becoming necessary to create genuine evaluation satellite posts.

A system of correspondents would entail a gradual transfer of methodology and know-how so that in the future regional centres would be able to carry out their own education performance surveys and construct indicators for measuring schooling efficiency. Regular evaluations using standard procedures and devised by local teams may be imagined, which would then serve as a national reference criterion. Change could be measured by systematic use of indicators of this kind.

Three sectors of student evaluation

With the passage of time, the evaluation process has ramified: it now extends in three directions:

i) The original evaluation process involves the *permanent monitoring of student learning* at key points in their school career. Sample classes are given standard tests designed to measure the achievement of objectives over a wide area. In this way, the monitoring exercise provides a set of indicators on efficiency and performance which enable actors, partners and users to be better informed.

ii) The second direction is *evaluation of education policy* to check the effectiveness of measures dealing mostly with pupils having problems with conventional schooling. By supplying those in charge with information that can enhance the quality of their decisions, evaluation acts within the system as an internal regulator.

iii) A recent development is an *"evaluation tool"* made available to teachers for assessing a pupil's strengths and weaknesses at the start of the school year and enabling teachers to readjust teaching methods accordingly. Evaluation of this kind sharpens teaching effectiveness and contributes to improved school performance.

The monitoring of student achievement

At the beginning of the 1980s, a permanent system for the monitoring of student learning was set up in France with a view to regularly checking the education system's success in teaching. A scheme covering pupils, teachers and schools was gradually introduced.

About every four years, surveys keyed to the decisive moments in the curriculum are conducted in order, on the one hand, to verify the performance of nationally representative samples of pupils and, on the other, to obtain factual information and comments from pupils and adults (principals, teachers), which shed light on how the system is functioning.

The monitoring system is now capable of measuring trends, and its database is an aid in constructing indicators for measuring student learning.

Scope of investigation

Annex 2 shows the scope of the monitoring exercise; it is very extensive, encompassing what pupils know and practise not only in respect of syllabus objectives but also outside the academic sphere. Thanks to specific questionnaires, the collected data cover cognition as well as know-how, attitudes and opinions.

Given the importance of proficiency in the "three R's" for later scholastic success, evaluation gives a special place to reading, writing and mathematics. But there has been a change of approach, especially where reading is concerned. In evaluations since 1987, for example, test papers have regarded reading as a tool for gaining access to other knowledge. Success in secondary school requires proficiency in understanding written material of various kinds. Evaluation must therefore use a wide variety of texts which include numerical tables, historical, documentary, practical and literary writings.

Evaluation is carried out in more specific subject areas such as modern languages, history, economics, technology, science, etc., using a variety of mediums (charts, tables, text, illustrations, audio cassettes, etc.) and requiring a similar variety of responses (multiple-choice questions, open questions, etc.). Evaluation deals not only with cognitive aspects, which are the easiest to test, but also with other, less measurable areas of know-how and method, such as the experimental approach to science.

The survey principle also opens up the possibility of testing at any given moment pupils' knowledge of a subject which is not yet contained in their syllabus. The information derived from these studies serves to indicate the ``zero setting'' for future schooling. Teachers can adjust course content and teaching techniques to the already acquired mental structures of the pupil majority. When the zero setting is known, they can more easily help pupils to attain a more complex understanding, in economics for example, and can better incorporate the elements of the subject into their courses.

Indicators of the degree of knowledge and know-how acquisition among pupils are, however, insufficient to give a picture of how a class functions. Indicators covering essential features of school life are also required. These have been worked out for lower and upper secondary schools and are interpreted from the replies of pupils, principals and teachers to special questionnaires that usually deal with three broad topics. The first concerns fitting into school life and endeavours to analyse a pupil's experience – with the *collège*, for example, being viewed as a community and as a formative environment. The second topic focuses on adjustment to community conditions and awareness of the outside world; it examines such issues as responsiveness to community living (joining in groups), initiative and leadership. The third is more directly centred on preparation for the future, *i.e.* the information sought, received and absorbed by pupils; the development of career ambitions; the effectiveness of pupil counselling and observation. The fact that the questionnaires submitted to the three sets of actors contain a shared body of questions means that comparative indices, useful for appraising the efficiency of the system, can be established.

Collection of additional information

All the evaluations carried out by the monitoring system attempt to gather details on context, pupil-related data and the opinion of teachers on their methods.

A description of the schools in the random samples is generally provided by the detailed enquiries conducted by the DEPA's Sub-Directorate for Studies and Statistical Surveys. The description lists the features of the rural-urban milieu, size of school, establishment structure and facilities. These data can be used for classifying the *milieu* of schools; for example multi-grade teaching or school-based innovations.

When information on a given class is collected, a fiche on each pupil surveyed contains details of age, sex, nationality, socio-economic status, parents' education, school antecedents, modern language(s) studied, expected educational career, etc. Cross-referencing these data with school performance data allows for the classifying of schools according to a taxonomy of types.

Questionnaires issued to teachers concerning their subject obtain their opinion on the importance of the objectives surveyed compared with the full range of objectives for a particular grade and on the difficulty of the survey tests administered to pupils, as well as their predictions of success before the tests are made. A short account of teaching methods tells about the application of the syllabus and, more especially, whether an idea has been studied before the survey was held. The monitoring system thus supplies decision-makers with a mass of information on the value of an idea in the school programme. It should thus be possible to decide whether to continue teaching the idea in the same grade, to postpone it to a later grade or surround it with additional instructions reminding teachers that it needs introducing or consolidating.

Assessment of relationships

The data gathering system is designed for making macroscopic observations. There is no possibility of deducing the performance of a particular pupil or school, since all the items in the sample are merged into a national average. The recorded observations are varied and somewhat complex:

a) The first observation concerns the proportion of pupils in a given grade meeting the standard under survey, in regard to a particular subject. The *coding* system employed gives the percentage of correct, wrong and non-replies, as well as the proportion of incomplete replies and common mistakes that were pinpointed during the pilot tests of the survey forms.

Coding should not be confused with marking, and it does not necessarily say anything about quality, beyond giving an accurate count of correct and incorrect answers. It expresses the pupil's reply as a symbol. It is an essential part of the method, and is not a mere technical device enabling the data to be computerised. The exact classification of the pupils' replies is closely correlated with the purpose of the question, so that the coding of the answers reveals the extent to which the objective has been achieved.

b) The findings can point to relations between items of knowledge and know-how within a given subject, as well as between several subjects. It is interesting to go beyond the results achieved in a single subject and discover the success pattern across a number of subjects.

A survey conducted along these lines during a tenth grade evaluation revealed the success patterns of pupils in five subjects at the end of the first upper secondary "decision" year. These patterns were based on scores in each of the test subjects judged to be higher or lower than the median (the benchmark at which 50 per cent of pupils have a better score and 50 per cent a poorer one). The pattern MfaPH, for example, would signify a pupil doing better than the sample median in mathematics (M), physical sciences (P), and history and geography (H), but less well than the median in French (f) and English (a). Upon completion of the survey, the findings showing the links between the subjects were as follows: a quarter of the pupils had an equivalent level in all subjects, whereas three-quarters had more varying levels – although it was rare to find contrasts within science subjects or literary subjects. English was the subject where the largest proportion of pupils was strong while being weak in the other four subjects. Conversely, this proportion was smallest in the physical sciences. French was the subject where the largest proportion of pupils was weak, while being strong in the others. This proportion was smallest in mathematics (ministère de l'Éducation nationale, 1989).

c) The findings also show a link between scholastic performance and other variables. Cross-matching between academic success, education system factors (school life, repeating, streaming decision, size of school, etc.) and personal details (age, sex, prior school career, socio-occupational category of parents, etc.) enriches the findings. With reference to the ambitious target of getting three-quarters of an age group as far as twelfth grade (*terminale*), it would seem that special attention should be paid to the streaming of pupils in upper secondary education.

During the tenth grade survey (which is a decision threshold after which the pupil must make a choice), a distribution of pupils by success pattern was compiled. Study of this distribution by stream, as it concerned the uniform patterns, produced the following statistics:

- Uniformly strong pattern MFAPH (15.8 per cent of pupils). 57 per cent of the pupils in this category were streamed in their eleventh grade into a science option (the élite S section), as compared with 23 per cent of the general school population. Some pupils (2.8 per cent) with this pattern were, however, streamed into a predominantly technical option whose candidates are typically miscellaneous and which generally leads to a "short" cycle of further studies, or were obliged to repeat their tenth grade (5.1 per cent). In the light of these figures, it is fair to cast doubt on the streaming standards practised by the school system.
- Uniformly weak pattern mfaph (9.9 per cent of pupils). Most of these pupils are directed to vocational colleges (18per cent as compared with 5 per cent of the general school population), streamed into eleventh grade technical (24 per cent as against 14 per cent) or required to repeat (29 per cent as against 16 per cent). 13 per cent of pupils showing this weak pattern are nevertheless oriented towards a "long" study cycle section in their eleventh grade. This is borne out by a school life question to which 84 per cent of teachers replied that class meetings try to offer a chance to as many pupils as possible rather than restrict the eleventh grade to pupils certain to succeed.

The processing of individual pupil data can provide answers to other questions raised within the school system. For example, is there a difference in academic performance between male and female pupils? Do they follow the same school careers? Do they feel the same about school life? Multivariate analyses can supply partial answers here. At the end of seventh grade, for instance, girls have better average marks, seem to fit better into the school, and tend to continue with a general studies career, more often than boys.

d) Findings enable the progress of knowledge acquisition to be charted among school populations. The data system can carry out diachronic follow-ups for measuring developments over a period of time. Recent seventh and ninth grade surveys were designed to illustrate scholastic performance trends several years apart. Setting the same tests can shed light on the controversy over "standards".

The system can also measure the progress in knowledge acquisition at several stages of the curriculum. The methods used involve either following a particular cohort of pupils from, say, their entrance in the sixth grade through to the end of lower secondary school or comparing statistically compatible samples of school populations. When the survey on the end of seventh grade was conducted in 1988, one aim was to measure changes and similarities with regard to the findings on the same grade in 1982, but another was to assess trends in learning with respect to certain acquisitions and difficulties, particularly with reading, observed in 1987 at the end of primary school (ministère de l'Éducation nationale, 1990).

The evidence on acquisition progress showed that understanding of reading material (based on six texts already tested on fifth grade pupils) increased significantly where the lower secondary pupils had acquired improved knowledge and where rearranging or

BEST COPY AVAILABLE

mentally rephrasing text content was required. Whereas individual items of information seemed well understood or well registered (as was already the case at the end of primary school), more profound understanding still seemed to be a problem. It is possible to see a relation here with the difficulty experienced by pupils in mastering the logical bridge words (such as "also", "moreover", "because", etc.) that are so important an element in mastering logical reasoning. Thus pupils at the end of seventh grade, despite their improved knowledge, still lack an adaptable method applicable to a wide variety of reading situations.

e) The findings provide information on the problem areas between the different layers of the education system. The abruptness of the transition from primary to lower secondary and from lower secondary to upper secondary school is much criticised. It is true that the new framework legislation posits unbroken education, but surveys from a few years ago show that there is a hiatus between, for example, primary and lower secondary school.

An identical test based on the final primary school curriculum was administered to a representative sample of pupils a few days after their induction into sixth grade and to other pupils finishing primary school. Identical questionnaires were also addressed to fifth grade and sixth grade teachers asking their opinion on the importance of the survey objectives and on their pupils' academic prospects. The general trends showed, first, that fifth grade schoolmasters expect more than sixth grade teachers. Whereas there was relative agreement on the usefulness for later schooling of the skills tested in the survey, the rate of undecided answers was higher among the lower secondary than the primary school teachers, the former holding more varied opinions than their primary school counterparts and predicting less bright prospects for their pupils than the results actually achieved. Second, academic performance was consistently higher at the end of primary school, especially in areas where school exercises were important (spelling, grammar, etc.); holidays had had a damaging effect.

There was a serious decline in performance in mathematics at the beginning of sixth grade, the widest variations occurring in respect of concepts that had already posed problems for at least a quarter of fifth grade pupils: calculating a surface area (39-27 per cent error), or the four sides of a rectangle (21 per cent error), multiplication and some division (19-16 per cent error). In French, the most frequent weaknesses involved comprehension, certain kinds of syntax, spelling – schoolwork still being learnt at the end of primary school.

Need for improvement

While the general approach of the survey design, especially the phase that must be devoted to defining objectives, can be applied to all operations, certain surveys require particular procedures. They fall into two categories: either information is not directly gathered and existing material is evaluated; or the subjects are difficult to measure quantitatively and experimental tools have to be devised and made reliable.

a) Evaluation using examination work: savings along with drawbacks. Learning target evaluation has been carried out in twelfth grade subjects such as history and geography. It was based on a sampling of *baccalauréat* examination scripts, so as to avoid a form of testing that would have been too much work for pupils and teachers.

The method consisted in studying official syllabuses and instructions on the teaching of these subjects in twelfth grade and the *baccalauréat* examination rules, using findings to draw up the list of objectives. The national steering committee then drew up an evaluation chart for the standardized analysis of the examination scripts. The chart was extensively tested beforehand to remove over-subjective criteria or terms over which the markers were deeply divided: a batch of scripts was given to evaluators to see how well each part of the chart was filled in, and items which were an obvious source of disagreement were discarded or reworded in the final chart.

It takes an enormous amount of work to extract the findings from this sort of exercise; *post factum* identification of examination objectives is difficult. Despite its limits, however, this method does give an idea of the implicit aims. The DEPA used a protocol of this kind when evaluating the papers of teachers (not pupils) sitting for the internal Secondary Education Aptitude Certificate (CAPES), a diploma needed to become a tenured secondary school teacher.

b) Some areas remain practically unexplored by evaluation. The systematic development of test procedures makes simplification important. Collective testing relies largely on "pencil and paper" exercises, sometimes supplemented by personal observation by the teacher using a check-list (*e.g.* noting laboratory work, physical education and sports activities, etc.). In the case of oral proficiency in a modern language (the ability to introduce oneself, make a request, transmit information), it can be awkward devising standards of appreciation. Similarly, evaluation of corporal expression touches on a highly complicated area where three factors (fluidity of movement, suppleness, originality) have to be assessed.

In testing talent for corporal expression, a sample of 220 pupils was given the following exercise – you have a hoop, an area bounded by two lines three metres apart, and exactly two minutes thirty seconds; in that time, create as many movements as you can going from one line to the other. The number of movements accomplished (fluidity) was used to situate pupils on a scale going from one movement (five pupils) to 37 movements (one pupil). To judge suppleness, *i.e.* the number of different categories of movement accomplished, a classification was made according to the position of the hoop (touching the ground, off the ground or in contact with the body) and the way the pupil used it (holding it with one or two hands, the foot or another part of the body, throwing it, rolling it, etc.). It was observed that five pupils used only one category, whereas one pupil used all 15 categories registered. 50 per cent of pupils used four or five categories. Originality was assessed according to performance and distribution across the 15 categories. Four headings were established afterwards: original (categories containing 0 to 5 per cent of performances), fairly original (categories containing 5 to 9 per cent), not very original (9 to 21 per cent), not at all original (over 21 per cent). As suspected, originality is a rare quality – only four pupils did better than the theoretical average.

This evaluation of originality in corporal expression revealed only a small number of performances falling into different categories. Some performances were spontaneous manipulations of the hoop as pupils learned how to use it. Much more rarely, complex patterns of movement and unusual handling of the hoop gave evidence of originality among the pupils.

Research is essential if evaluation is to progress. No matter what the field – written work, oral expression, pupil attitudes (teamwork, etc.) or general abilities common to all subjects – the ideas and methods of university research teams should help in the development of experimental tools for measuring complex observations.

Interdependence of monitoring and policy evaluation

Alongside collecting information and providing a record of pupils' attainment of official targets, the intermittent evaluation of education policy exercises a regulatory effect on the system as a whole. This type of evaluation, which attempts to gauge the likely effects of new measures, pinpoints any discrepancies that may exist between intentions and reality. It gives the authorities valuable information for quickly reviewing a measure and improving its potential educational impact. As the Planning Commissioner, Pierre-Yves Cosse (1989) pointed out, "You cannot make headway unless you know your position at the start and the effects of the measures you undertake. This requires an evaluation of both pupils and schools."

As an illustration of the above remarks, this section proposes dealing with two kinds of evaluation of education policies designed to help pupils in difficulty:

- The first is an evaluation of the eighth and ninth technology grades created for pupils experiencing difficulty with traditional schooling. Thanks to courses tailored to the role of technology, these classes open up a new avenue of possibilities for pursuing further studies.
- The second, still in progress, is an evaluation of all the measures taken to assist underachievers entering the observation stage (sixth and seventh grades) of lower secondary school. Schools have in the last few years instituted two sorts of programme for aiding these pupils – a three-year instead of a two-year period for covering the syllabus, and more individual tuition techniques such as remedial classes, catch-up groups, tutoring, supervised studies – both expensive and in need of cost-effectiveness assessment.

Measuring the effects of education policy

Any new policy measure or innovation intended to improve the the education system should be combined with the use of survey instruments for evaluating the consequences of such measures. Each new measure usually comprises a statement of aims, an account of the problem involved and quite often a summary of the hoped-for results. These elements are of use in the evaluation process.

Not everyone agrees with evaluating education policy. Some people criticise the over-hasty evaluation of structures and teaching practices that have not fully adjusted to the introduction of a new measure. They advocate a cautious approach and advance the following questions as arguments: Is it not preferable to wait until the new measure is more fully integrated into the system and more widely applied? Are not the findings liable to give only a rough picture of a shifting reality? The answer is "No". It is appropriate to ascertain the situation in the initial stages of a new policy and undertake a further evaluation some years later in order to measure any changes due to the wider application of the measure. Policy-makers needs effective tools which will help them to make good, informed decisions during the implementation phases of new measures. In the light of the information supplied, decisions to drop, reinforce or alter the measure will have been carefully weighed.

Recording the circumstances of policy implementation

Who decides to implement new education policies or innovations? Are they set up and applied within a school project involving the whole staff, or are they ordered by the principal? How and at what stage are they carried out? Are they applied in a consistent manner?

In the case of the technology grades previously mentioned, considering that they are being expanded at a fast rate (12 054 pupils enrolled in eighth technology grade in 1985, 61 588 in 1989), there was a striking difference in the way they were established, according to whether they were introduced in a vocational upper secondary school preparing the "short" cycle or in a general education college. In the vocational schools, they usually replaced arrangements catering for pupils in serious difficulty; in the colleges, they tended to be new structures. As may be imagined, these differences induced very dissimilar styles of pupil enrolment.

It is important to document how pupils are affected by an education policy measure, for example: by considering characteristics such as age, previous school career, socio-occupational group, French-born or foreign? Concerning the application of the policy measure, have teaching groups been formed with teachers who have received special training and act as a team in framing a concerted project?

Information gleaned from questionnaires submitted to principals, teachers and pupils offers the institution a picture of how a measure is being implemented, and indicates whether, and to what extent, it is meeting expectations.

Measuring attitudes, expectations and satisfaction

Submitting questionnaires to pupils and teachers is a way of obtaining an exact idea of the relational climate surrounding the introduction of a new measure – pupil receptiveness, more personal contact with teachers, requests for help from teachers when difficulties arise, pupils' perception of their own school performance, changes noticed in teaching practices, etc.

In-depth studies give shape to the data collected (*e.g.* factor analysis, construction of indicators) and allow indices to be compiled (*e.g.* index of conformity with official instructions). In the survey of the technical grades, analysis of replies revealed sub-sets of teachers who were more familiar than others with the instructions, believed more in the rationale of these classes, because they met more often to exchange ideas and carry out a teaching project, and were interested in motivating underachieving pupils. Concordance between this information and data supplied by the pupils showed that the same classes were in compliance with the official directives; this is an indication of how the new teaching practices are taking hold (Hornemann, 1989).

Comparing achievement with stated goals

The sophisticated longitudinal study carried out in connection with the survey of student performance in the sixth and seventh grades highlights the problems involved in evaluating education policy.

This kind of study makes it possible to answer a number of questions arising during its period of operation, since the initial and final situations are known. What are the grounds for assigning a pupil to a three-year cycle instead of the usual two-year one? Is the diagnosis just? After the three-year cycle are pupils as proficient as those who have followed the two-year cycle? Are there differences between school populations according to age, sex, nationality, socio-economic group, etc.? Does the lengthened cycle produce the same results as a repeated year? Do the pupils from the lengthened cycle continue their studies in the same way as pupils from the ordinary cycle, in entering a general or a technical eighth grade?

The 1988 evaluation of the technology grades, once completed, showed uneven achievement of the goals set for these classes, and the policy-makers were duly alerted. An encouraging finding, by contrast, was that pupils who had been in difficulty in the normal system regained confidence. Another bright point was the orientation of a maximum number of pupils towards a high-level technical diploma, in line with the wishes of parents and pupils, who found themselves reconciled with the school system. The findings were much less clear-cut, however, concerning teacher teamwork in developing a teacher-pupil effort to enhance abilities and promote the acquisition of interdisciplinary skills. Similarly, the use of technology seems to have deviated from its initial purpose, which was to develop the capability for pursuing further studies. Technical education, especially in the classes set up in the vocational high schools, seems to lead to job specialisation.

As for the remedial classes, the results are very inconsistent; after finishing their cycle, a large majority of pupils show gaps in their basic knowledge and capacity for logical reasoning. What will happen to pupils who cannot graduate to higher classes or obtain an occupational diploma? Since the creation of the technology grades has gone hand in hand with the abandonment of the old job qualification system (occupational aptitude certificate), these pupils might well swell the ranks of young people leaving school without qualifications. One of the advantages of education policy evaluation is to call the education system's attention to the problem, which it is incidentally trying to palliate. The findings of the latest survey in 1990, currently being processed, should confirm the indications of the 1988 enquiry, or they may alter them now that the situation is more settled and the partners have presumably become more familiar with the directives.

If education policy evaluation is to be effective, it must go beyond simple observation. While it need not spell out complete practical recommendations, it should always include a set of proposals on which the education system may focus in trying to improve the measure in question. Then, once the innovation or policy measure has become a widely accepted part of the system, it will be dealt with by the monitoring system described previously, and subjected to routine forms of evaluation.

Improved understanding of the system should improve learning

One of the priorities of French education policy is to improve the success rate of primary schools so that most pupils may continue their schooling without major difficulty, and so that Eighty per cent of pupils in a given age group will stay on until the end of secondary school by the year 2000.

As Lesourne (1987) points out, "Seeing that primary education leaves an indelible mark on children, should not the priority of priorities over the next decade be to improve its efficiency and performance? The first task is surely to upgrade the quality of primary teaching since it is the rock supporting the whole edifice."

Ninety-five per cent of pupils continue into lower secondary school, but surveys have shown that, whereas 85 per cent of pupils can read at the end of primary school – that is, they can decipher and extract information from a text – a little more than half only are good enough at reading to use it as a tool for acquiring knowledge independently. This being so, the ambitious goal for the year 2000 has no hope of being attained unless appreciable progress is made in teaching and learning the basics.

Following a study by Migeon (1989), the Minister of State, in February 1990, decided on a number of important measures for a new primary school policy. These measures emphasized a new "pupil-centred" teaching approach:

> "The pupil, accepted in his or her psychological, physiological and social reality, must everywhere become the central criterion for education. While maintaining its standards of quality, our education system must readjust its teaching techniques so as to bring its objectives within the reach of every pupil and thus restore the credibility of those standards. This is where the difference lies between a system which too often does no more than punish failure, and one which helps the pupil learn by his or her mistakes." (ministère de l'Éducation nationale, 1990, February)

As part of this policy, teachers had to be given the means to evaluate their pupils accurately and quickly. At the beginning of the 1989 school year, a survey operation was undertaken as a teaching tool for individual pupil assessment. It comprised three elements: evaluation, briefing, and feedback to pupils. These three parts of the operation are interlinked. Evaluation of pupils reveals the difficulties they are encountering; the briefing of teachers depends on the findings and helps with modifying teaching practice.

National evaluation for increasing classroom effectiveness

The evaluation protocol employed in the survey operation mentioned above differs from the other surveys carried out as part of the monitoring system, which are aimed at collecting general information about the education system. It is designed, first and foremost, to give each teacher a classroom instrument for rapidly and thoroughly diagnosing pupils.

There is another difference compared with the more conventional surveys: the tests are not designed to situate pupils on a scale of performance; they are intended to help identify the children showing basic learning deficiencies at the start of an education cycle and determine the nature of the difficulty encountered.

Since the aim of the exercise is to improve teaching efficiency in the classroom, it may be asked how the teacher's role is modified by this new tool. According to how interested they are in using instruments of this kind, some teachers seem to interpret the results superficially, saying that they merely confirm what they already knew about their pupils, while others make a more careful analysis of their pupils' difficulties and draw a parallel between the kind of difficulties noted both in mathematics and French. Others again explored the findings as an effective technique for diagnosing their pupils' characteristics right at the start of the school year and spotting weaknesses that would need remedying. It was discovered, for example, that some pupils did not understand the purpose of writing, did not grasp the connection between speaking and writing and were unfamiliar with written material coming in various forms. This observation led teachers to diagnose the real reasons for failure and allowed them to take remedial action such as varying the applications of writing, getting their pupils to "do their own homework" on different types of writing, etc.

The survey also gives the teachers the opportunity to draw up a table of class results, which can be used in adjusting teaching strategies towards pupils or groups of pupils. A number of reports show that teachers have used the survey to pick out children requiring remedial aid in reading. This enabled them, very early in the new school year, to form groups of pupils who would benefit from individual attention.

Since evaluation is personal, it can act as a basis for contact with parents. Each teacher can invite families to come and discuss their child's results. A free half-day is allowed for this consultation. Eighty per cent of parents said they had welcomed this early opportunity for dialogue with the school; such a percentage of participation is unprecedented.

Reports also indicate that the survey is valuable for making changes in teaching practice: reorganising class activities (giving individual attention to pupils, short-term increases in the time allotted to class study of certain subjects, etc.) and school routine (*e.g.*, assembling pupils from different classes for remedial tuition in particular areas of study). The survey can also stimulate teamwork with teachers of the grades above and below the ones surveyed.

Using the survey results, teachers can judge the effectiveness of their approach, taking into account the kind of pupil in the class concerned and situating pupil performance in comparison with national standards derived from an aggregation of the total scores of the survey. Interested schools have been supplied with score calculation software to help them recap findings. Because of the heavy demand, consideration is being given to devising more powerful software for helping teachers to make fuller use of the data obtained on their pupils' attainment.

National evaluation and system management

When evaluation has been conducted over an extended period of time, it can offer valuable information for orienting and managing the education system. It can provide decision-makers with reliable information for gauging the effectiveness, in the medium-term particularly, of measures concerning the effects of training schemes. It gives local, regional and national education authorities evidence of the achievement of goals in areas with high percentages of underachieving pupils. The data can provide justification for increased budget allocations or special action to remedy underperformance.

National evaluation and school efficiency

In addition to the individual results transmitted to parents, the national survey findings are widely circulated nationally, regionally and among schools. Each teacher taking part in the survey is sent a booklet containing the list of results item by item, together with sub-scores and the overall average score, as well as comments on the major trends observed. The results have received a great deal of media attention, showing the interest of the public in the quality of education.

As soon as the first national survey was announced, the media gave a detailed account of this initiative to improve schoolchildren's chances of success. When the findings were published, numerous articles appeared in the national and regional papers, the educational press and the journals of the teachers' associations. The main subjects treated were the novelty of the measurement (high-quality tools and methods), the enthusiasm of parents, the disgruntlement of the teachers faced with the heavy new workload, and the teachers' demand for training. More attention was paid to the shortcomings of the children's performance than to their attainments. The initiative is widely considered a success: a vast majority of teachers and families understands and supports the aims of this action conducted on behalf of the children.

Conclusion

The preceding sections dealt extensively with the question as to how pupil knowledge acquisition surveys can be exploited. Evaluation, however, deserves treatment in its own right, since it is a driving force behind decisions adopted by the actors in the education system.

Even where its effect is not immediately discernible, evaluation creates a climate favourable to decision-making – and this is true despite the existence of problems or of differences between the applications predicted by the evaluators and what actually occurs.

In conclusion, the possible applications and actual uses made of evaluation and assessment studies in France are summarised below.

Scope of information

Decision-makers determine or alter policies in an effort to perfect the education system. To do this they need performance indicators, particularly on:

i) *Scholastic performance*

With this knowledge, teachers can adjust their methods to improve the match between objectives and attainment. Syllabus designers can rethink teaching content in terms of its coherence and gear it to pupils' abilities.

ii) *Differences in scholastic achievement*

Evaluation is useful for understanding pupil strengths and weaknesses; it not only reveals the proficiency or lack of proficiency of the average pupil, it also reveals the standards of the strongest and weakest pupils. This knowledge provides teachers with a basis for helping weak pupils.

iii) *The particular characteristics of a system*

The French education system has specific features whose effectiveness requires evaluation: the benefits of repeating, the criteria applied in streaming, the usefulness of pre-schooling, etc.

iv) *The changing state of learning acquisition*

Comparison, whether of diachronic trends or learning developments in a curriculum, is a vital element in the management of the system. International comparisons will provide even more data on aspects of the education system that may be in need of reform.

The various surveys carried out at primary school level, have highlighted the problems experienced by many pupils in using reading as a way to acquire knowledge, especially when it comes to reading the terms of a problem, as in mathematics. These important observations, in the context of the plan to keep a high number of pupils in the system until the end of secondary school, have done much to influence primary school reform.

It is worth mentioning that the central goal of this organisational reform – to pay greater attention to the individual child's development – relies on evaluation. Another important point is that the practical effect of the survey findings depends a great deal on how they are presented and disseminated.

Effects on teachers

Each survey begins with a choice of goals, the creation of standard tools and the definition of objective criteria for evaluation. Now that the monitoring system has become fully operational and exhaustive national surveys are being conducted, curriculum planners and teachers who take part in evaluation surveys are exposed to many new ideas. Participation is thus a form of in-service training; for example, teachers learn:

i) about the choice of goals;
ii) about evaluation methodology;
iii) about evaluation criteria;
iv) about the factors influencing attainment.

Care must be taken if the real benefits of national evaluation are to be lasting. A new approach to teaching calls above all for teacher training initiatives based on the concrete findings of the surveys. Since change depends on shared convictions, teachers need to be involved in the various stages of evaluation so that they can gradually master an instrument which will allow them to gauge the effectiveness of their own teaching.

Transparency

The education system itself must supply users with unbiased information on its performance. If it does not, others will attempt to use any available information for their own ends. It is essential to aim at transparency; this will provide a better understanding of how the system operates and contribute to more dispassionate debate on the quality of education.

There is a problem, however, in striving for transparency when the findings of a survey are unwelcome. Should shortcomings be swept under the carpet in order not to spark public controversy or discourage the teaching profession, even though sooner or later the system itself may have to provide an accounting? It can be riskful to move from in-house communication for professionals to information for users and the public, yet this is necessary in an open system of schooling catering for very large numbers.

Teachers need signposts, as do principals, for assessing their own work. In the case of the primary school evaluation surveys, it is fairly certain that the national findings will be published, but much less so that the regional results will be divulged, for fear of invidious comparisons. Yet most people are aware that differences exist; should they be hushed up? One of the current features of French education is its great variety in terms of pupils, abilities, schools and teachers. This variety needs publicising to dissipate the monolithic image that prevails in the minds of many parents and teachers.

It is important also to provide parents – within the limits of the regulations and school districting requirements – with the information they need to choose the school where their child will pursue his secondary studies. Parents' attention should also be drawn to the unstable nature of pass-rates, which seem to fluctuate from year to year, and the highly selective character of some schools which can impose heavy demands on pupils with disastrous psychological results.

Evaluation standards

Findings must not be considered suspect merely because they do not match prior expectations. Pupil knowledge acquisition indicators are compiled to help decision-makers improve the overall performance of the education system.

Measuring school achievement over a period of time is a recent development, yet the latest outcry over the stagnation (or decline) in success rates in certain subject areas after two years in lower secondary school raises the delicate issue of how far survey findings are accepted.

A mechanical exploitation of surveys, which loses sight of their deeper purpose, must also be avoided. The danger lies in using them normatively, *i.e.* as a disguised exit examination at the end of a cycle – pass your evaluation test or else we will not let you enter the next class. This would lead to questionable practices among teachers as well – teaching only the survey education targets and coaching with a view to answering the survey test papers. Evaluation for evaluation's sake would become a sterile exercise unable to contribute to improving the system.

Evaluation was not originally a response to an outside demand; it represented an attempt to understand and measure the efficiency of education. Additional evaluation schemes are needed to complement or corroborate information from every stage of the system, for example, surveys conducted by outside bodies (polls, monographs, studies, etc.) or by the pupils themselves.

Proliferation and usefulness of evaluation

Surveys, as has been seen, differ in their aims and content. Pupils' knowledge is surveyed in connection, for example, with current curricula but it is also assessed along non-school lines. Other enquiries deal with know-how, methods, cross-disciplinary abilities, opinions, etc. Further variety has been provided by testing innovations for improving scholastic achievement, or providing teachers with evaluation tools. The proliferation of evaluation makes it difficult, however, to assess its immediate or long-term usefulness.

The national evaluation of pupil knowledge acquisition is one source of information among others. It may not always lead to tangible changes, but it does add something to the "backdrop" against which decisions are taken.

Evaluation is one element that makes decision-making easier. Its reliable observations can help authorities to reach a sound decision, rather than one taken "in the dark". Evaluation nevertheless occupies an uncomfortable position at the point where short- or medium-term policy demands and the longer-term interests of the education system intersect. It can reconcile the two and assist in improving the system's guidance and regulation by ensuring that its goals and methods remain consistent.

Evaluation is no rigid formality, since it is the dynamic support of an approach by which goals are set with a view to action. As for its usefulness and the need to develop evaluation of pupils' learning acquisition, the reader may reflect on the issues raised in the subsequent chapters.

BEST COPY AVAILABLE

Annex 9.1 **Comparison of the international standard to the French primary/secondary**

		Age	US grade equivalent	
ISCED 0		2		
		3		pre-primary education
	middle section	4		
	upper section	5		
ISCED 1	CP	6	1st	primary education
	CE1	7	2nd	
	CE2	8	3rd	
	CM1	9	4th	consolidation, elaboration
	CM2	10	5th	
ISCED 2	6th	11	6th	
	5th	12	7th	observation
	4th	13	8th	
	3rd	14	9th	orientation
ISCED 3	2nd	15	10th	decision
	1st	16	11th	
	Final	17	12th	

Annex 9.2

Type of school	Grade evaluated and date	Subjects evaluated
Primary school	CP 1979	Reading - Mathematics
	CE2 1981	French - Mathematics
	CM1 1981	French - Mathematics
	CM2 1983	French - Mathematics - General cultural subjects
	1987	Reading - Mathematics
	CE1 1988	Reading - Writing
	1990	Reading - Writing - Mathematics
	CM2 1990	Reading - Writing - Mathematics
Lower secondary school (Collège)	6th 1980	French - Mathematics
	5th 1982	Same + English - German - Physical sciences - Natural sciences - School life
	1988	French - Mathematics - School life
	3rd 1984	French - Mathematics - English - German - Experimental sciences - History and geography - School life
	1988	Notions of economics
	1990	French - Mathematics - English - German Cross-disciplinary test - School life
	3rd technical 1990	French - Mathematics - English - Technology (Industrial science and techniques - Services - Biological and social services)
Upper secondary school (Lycée)	2nd 1986	French - Mathematics - English - German - History and geography - Economics and social sciences - Physical sciences - Natural sciences - Technology - Physical education and sports - School life
	1st 1987	French
	Final (Terminale) 1987	History and geography (using baccalauréat scripts)
	1988	English (using baccalauréat scripts)
	1989	Notions of economics (in general, technical and vocational streams)
	1990	English (oral and written comprehesion, oral and written expression)

Reminder: CP = 1st Grade; CE1 = 2nd Grade; CE2 = 3rd Grade; CM = 4th Grade; CM2 = 5th Grade; 6th = 6th Grade; 5th = 7th Grade; 4th = 8th Grade; 3rd = 9th Grade; 2nd = 10th Grade; 1st = 11th Grade; Final = 12th Grade.

Bibliography

BAUDELOT, Ch. and ESTABLET, R. (1988), "Le niveau intellectuel des jeunes conscrits ne cesse de s'élever", *Éducation et Formations*, pp. 3-15.

BOISIVON, J.P. (1990), *Projet 2001 l'horizon de l'école*, Round Table on the theme "What school policy?", Paris, November.

COSSE, Y. (1989), *Une formation pour tous*, report by the Commission Éducation Formation Recherche, La Documentation française, Paris.

DUTHOIT, M. (1988), "L'enfant et l'école – aspects synthétiques du suivi d'un échantillon de 20 000 élèves des écoles", *Éducation et Formations*, 16, pp. 3-13.

HORNEMANN, J. (1989), "Pratiques pédagogiques, climat relationnel et motivation scolaire en classes de 4^{e} et 3^{e} technologiques", *Éducation et Formations*, No. 21, pp. 63-77.

LESOURNE, J. (1987), *Éducation et société demain – à la recherche des vraies questions*, report commissioned by the minister of Education, December, Éditions La Découverte et journal *le Monde*, Paris.

MIGEON, M. (1989), *La réussite à l'école – quelques propositions*, report to Monsieur L. Jospin, Minister of Education, Youth and Sports, January, ministère de l'Éducation nationale, Paris.

MINISTÈRE DE L'ÉDUCATION NATIONALE (1989), "Évaluation pédagogique en classe de seconde, avril 1989", *Dossiers Éducation et Formations*, Vol. 5, pp. 5-156, ministère de l'Éducation nationale, Paris.

MINISTÈRE DE L'ÉDUCATION NATIONALE (1990), "Évaluation pédagogique en fin de cinquième", *Dossiers Éducation et Formations*, Vol. 6, p. 128, ministère de l'Éducation nationale, Paris.

MONDON, P. and SEIBEL, C. (1987), *L'évaluation pédagogique: un outil essentiel de l'action éducative*, Association française des enseignants de français, Paris.

PROST, A. (ed.) (1983), *Les lycées et leurs études au seuil du XXI^{e} siècle*, report presented by the National Working Group on Upper Secondary Education chaired by A. Prost, November, ministère de l'Éducation nationale, Paris.

Chapter 10

Theories of Mental Development and Assessing Educational Outcomes

by

Ivan Ivic
Belgrade University, Yugoslavia

Indicators of educational outcomes are of central interest to the INES Project on international indicators of education systems. The reason seems obvious: indicators on outcomes can offer useful feedback for improving the process of education. But if one considers, on the one hand, the swift changes in the objectives and functions of education in modern societies (*e.g.* OECD, 1992) and the approaches commonly used in the assessment of educational outcomes, on the other, then a number of fundamental incongruities appear. The following are just a few examples:

- On the one hand, new objectives in education include the acquisition of complex competencies, such as problem-solving abilities, high-level thinking, critical and creative thinking, and the formation of complex social competencies such as co-operation, communication and team spirit. Students are also expected to become independent active learners. On the other hand, the dominant doctrines of assessment still bear the stamp of theoretical concepts and objectives that originated in a different socio-cultural context. They emphasize assessing knowledge in particular school subjects, end-state school knowledge, and reproductive types of knowledge. They do not generally attempt to assess the process of acquiring knowledge.
- Modern education strives to employ all human resources, *i.e.* to develop every individual pupil, whatever his ability (hence the notion of "quality education for all"). Yet the dominant doctrine for assessing individual student achievement, *i.e.* the diagnosis of individual differences, has traditionally played an important role in allocation – the selection of students for educational and occupational careers.
- A basic function of indicators is to offer information and feedback on the process of education. In contrast, the more traditional approach to assessing educational outcomes implied measuring some final state, without offering insight into how this final state was achieve.
- Although education is an extremely complex endeavour, most conventional approaches to student assessment cannot handle this complexity and must simplify, for example by having a single number that expresses an achievement test score.

It may be possible to find a means to escape this situation by defining complementary and/or alternative doctrines for assessing educational processes and outcomes. However, arriving at useful indicators in the operative sense, indicators that satisfy the new requirements, will be a long process. The first step is to attempt to define a new conceptual framework for new categories of indicators. The elements of this new framework should be sought in all the domains that deal with the nature, development, and use of human knowledge (epistemology, history of human knowledge, cognitive sciences, the psychogenesis of cognitive abilities, etc.). This contribution is based on the assumption that theories of mental development, and particularly theories of cognitive development, may serve to inspire the formulation of new conceptual bases for assessing educational processes and outcomes.

Scope and purpose of the chapter

Basically, theories of mental development try to explain learning behaviour at different stages of development. Specifically, they provide insight into how people behave when faced with situations in which they have to acquire, process, and use information about the physical and social environment and about themselves. As education deals both with the developing human being and with the acquisition of knowledge and skills, it is clear that theories of mental development and the empirical knowledge about cognitive development generated by them are very relevant for an understanding of the process of education and its outcomes.

This basic assumption underlies the analysis, presented in this chapter, of problems related to educational outcomes in the light of significant theories of mental development. The chapter will discuss:

- the theory of L.S. Vygotsky;
- the theory of J. Piaget;
- the theory of modern cognitive psychology, particularly developmental cognitive psychology;
- instructional theory (the theory of the psychology of instruction or cognitive instructional psychology).

The theories of Vygotsky and Piaget will be the primary focus; the theory of developmental cognitive psychology and instructional theory will only be used to elucidate points raised by the analyses of the theories proposed by Vygotsky and Piaget. There is no need to review the theories of Vygotsky and Piaget in any detail. Rather, components and aspects of these theories which seem relevant to assessing and interpreting educational outcomes have been selected for examination.

Theories of mental development in general – and hence those of Vygotsky and Piaget – are primarily concerned with problems of cognitive development and the abilities needed for creating, acquiring and using knowledge; they are not directly concerned with problems of school learning. Insights offered by these theories can be used only indirectly to solve the problems raised by the assessment of educational processes and outcomes. Complex extrapolation and adjustment of these insights would be needed to solve specific problems relating to the indicators of educational outcomes.

The application of theories of mental development can help close the gap between theoretical conceptualisations and basic knowledge about cognitive development, on the one hand, and the theory and practice of school teaching and learning, on the other. Basic research on cognitive development traditionally has had an experimental orientation, while research on the process of knowledge acquisition at school has not taken into consideration the development of cognitive abilities. It has tended rather to reduce learning to a process of acquiring information. Both approaches fail to recognise that confrontation with a structured body of knowledge mediated by the school represents a major intellectual and developmental task for all children.

What is the potential contribution of theories of mental development to elucidating the processes and outcomes of intellectual development through school learning and to obtaining adequate indicators of educational outcomes? The conceptual framework which this chapter proposes to define might serve several purposes:

- It can serve as a basis for the critical analysis of the existing models of assessment and previously defined indicators.
- It can point to mechanisms that link the various indicators (especially process and outcome indicators); in other words, it can shed light on the way indicators of educational outcomes are produced.
- In this manner, extrapolations from the theories of mental development can generate a framework for interpreting the existing process and outcome indicators and their interrelation.
- It can serve as a starting point for defining a new generation of indicators of educational processes and outcomes that would complement or be an alternative to the existing ones, and thereby increase the validity of indicators in societies in which the function of education has changed.

In the discussion that follows, aspects of Vygotsky's and Piaget's theories are examined in turn, in an attempt to deduce the implications of these theories for indicators of educational outcomes, using aspects of the theory of cognitive developmental psychology and instructional theory where necessary. In a concluding section, a list of implications for education indicators is presented, and certain generalisations and extrapolations are made.

Vygotsky and the assessment of educational outcomes

The components of Vygotsky's theory that are relevant to the assessment of educational processes and outcomes are:

- allomorphic development;
- the relationship between school learning and mental development;
- the origin and function of metacognitive abilities.

Vygotsky's theory about the formative role of social interaction and co-operative learning between adults and children unites these three components.

Allomorphic development

One of the novelties of Vygotsky's theory is his view of the relationship between the individual and the socio-cultural environment, a view associated with culturalist theories of development. The common characteristic of these theories is the great significance accorded to socio-cultural variables. However, there is a major difference between Vygotsky's theory and other culturalist theories. While most theories attribute a motivational (dynamogenic) function only to socio-cultural variables, Vygotsky's theory ascribes a formative (constructive) role to them. This means that, in their absence, higher mental functions cannot develop; socio-cultural factors thus determine higher mental functions (Vygotsky, 1978; Ivic, 1987, 1989).

Vygotsky's theory also introduces the concept of allomorphic development. This means that culture, besides having a formative role in the genesis of higher mental functions, creates, in the course of history, external tools – instruments, mechanisms, technical appliances, aids, technologies – which support specific mental functions. These external tools appear as extensions and amplifiers of the natural powers of mental functions.

Vygotsky expressed this idea on several occasions in terms of the Latin dictum (borrowed from Francis Bacon): *Nec manus nisi intellectus sibi permissus multam valent: instrumentum et auxilibus res perficitur* (both the human hand and mind, if deprived of instruments and aids, are quite powerless; these instruments and aids increase their power). What is specific to Vygotsky's theory is his introduction of the dimension of external support to mental functions into his conceptualisation of development. He also deduced the consequences of allomorphic development for the structure, functioning and efficiency of the human intellect. In order to understand the impact of this idea for individual development, one might imagine the process and effects of problem-solving by three persons: one from a non-literate culture, one using the aids offered by a literate culture, and one equipped with modern information technology (Eraut, 1989).

Modern information technology and its impact on human learning and problem-solving processes have highlighted just how avant-garde Vygotsky's theories were in the early 1930s. They opened up the study of change in the structure of mental functions, by encouraging the use of both external instruments and the internalised techniques of intellectual work. For example, the relation of memory to thinking has changed and will continue to change as the function of memory is taken over by external aids. The increasing significance of various technical and technological means in the modern learning process reveals the importance of Vygotsky's conceptualisation of allomorphic development and its impact on individual development. Some of the implications for the assessment of educational outcomes are:

- The existing systems of assessment, which deprive the test subjects of external supportive tools, should be subjected to critical analysis from the standpoint of ecological validity. If in the learning process a student uses tools and aids which differ from those used in real-world problem-solving, then depriving the student of these tools and aids in the course of assessment can decrease the prognostic validity of such assessment.
- The content to be assessed could be enriched by ideas about allomorphic development. For example, a significant component of assessment should be an assessment of an individual's ability to use various instruments and information sources

such as public computerised databases. The individual's ability to create and use personal databases should also be assessed.

- Assessment techniques that make use of paper-and-pencil techniques could be modified so as to allow students to use different auxiliary tools in the course of testing, thereby shifting the emphasis of assessment to the ability to use these aids. This in turn would mean that the ability to memorise would be appraised to a lesser extent, and thinking ability, *i.e.* problem-solving ability with the aid of various auxiliary tools, would be appraised to a greater extent.

Mental development and school learning

The part of Vygotsky's theory which specifies the link between mental development and school learning is of great significance. Vygotsky is among the few scholars to have seen the relationship between development and school learning as one of interdependence. In this view, both in terms of causality and in terms of moment of appearance, there is a connection between mental development and school learning. Not only does development bring about learning, learning also promotes development.

According to Vygotsky, the greatest intellectual real-world task of children in modern societies is to assimilate a structured body of knowledge during a long period of systematic school instruction. In referring to systematic school learning, Vygotsky spoke of "the development and formation of concepts" and not only of their acquisition. These he referred to as "real concepts" – as opposed to "experimental" ones, which are usually studied in psychological experiments. They should not be seen as isolated concepts but as "systems of scientific concepts". The acquisition of this structured knowledge is at the centre of cognitive development, and it should not be reduced to learning only, that is, to the acquisition of a body of "ready knowledge". Glaser has expressed the same idea, albeit in a different way: "intellectual functioning before, during and after specified instructional treatments is increasingly a central topic" (cited in Snow and Yalow, 1988, p. 497).

Essential to this view is the idea that the intellectual tasks the child is faced with are not concerned with isolated items of knowledge, or even isolated problems, but systems of concepts. Comparing the knowledge acquired through the child's personal experience and scientific knowledge, Vygotsky says: "The essential characteristic is the existence or lack of a system" (1956, p. 309). Other characteristics of conceptual knowledge – the specific relationship of concepts to the objects they refer to, the existence of a network of relationships between concepts of different hierarchical levels, and the intellectual operations possible inside the system of concepts – also follow from this basic trait.

Vygotsky has described in detail one of the possible structures of scientific knowledge: namely, the hierarchical organisation of concepts. He recognised that there are other kinds of organisational structures of concepts besides taxonomic hierarchies, including: genealogical trees; structures of knowledge in the form of algorithms or systems of algorithms (so-called procedural knowledge); systems of rules; systems of axioms; and rules for generation.

In the hierarchical organisation model, one can distinguish three aspects or levels. The first is manifest content, *i.e.* concrete facts, data or information specific to a domain of knowledge. This aspect is close to the knowledge termed declarative or conceptual knowledge in cognitive psychology. It is this level that is generally pointed to when it is argued that acquiring the content of school subjects is not important for the development of thinking. A great majority of classical tests of knowledge assess this content level.

The second aspect or level of the content of school subjects and the corresponding disciplines is the instrumental one. This includes all those components of knowledge in any discipline that concern methods, techniques, procedures, skills and technologies. Some are close to what has been described in modern cognitive psychology as procedural knowledge. This can be regarded as the operative side of knowledge, and it can be divided into two categories. One is general instrumental knowledge, which is not specific to a certain domain; it would include general techniques of information processing and general intellectual skills. The second is the instrumental knowledge that is inseparable from specific knowledge in a certain discipline, so that it can only be acquired at a certain level of expert knowledge in that discipline.

The third aspect or level of the content of school subjects is the structural one. This is the most abstract level of knowledge, and it contains modes of thinking that are specific to certain domains of knowledge. These modes of thinking include experimental thinking, axiomatic thinking, algorithmic thinking, probabilistic thinking, historical thinking, etc.

In sum, the knowledge that is mediated by the school contains intellectual tools and modes of thinking. The systems of knowledge and the models of thinking that underlie school knowledge therefore contain built-in intelligence.

The analyses of the epistemological nature of knowledge carried out by Vygotsky are close to what is called "task analysis" in modern cognitive psychology and cognitive instructional psychology. Task analysis is a basic concern of cognitive psychology, including developmental cognitive psychology. Research performed in the Piagetian tradition of developmental psychology can also have as its objective the discovery of the epistemological nature of certain intellectual tasks. For an analysis of the tasks of conservation and class inclusion, see Inhelder *et al.* (1974), and for the concept of time, see Montangero (1985). However, there is a difference between the analyses of Vygotsky and task analyses in cognitive psychology and Piagetian research. Where the latter prefer precise analyses of the nature of individual tasks, Vygotsky tries to analyse the organisation of knowledge.

The basic characteristics of the intellectual tasks that a child faces when acquiring a system of knowledge in the course of schooling are:

- Formation of individual concepts, a process that has been thoroughly studied in laboratory research under the title of "concept formation".
- Comprehension of the organisation of knowledge, including instrumental and structural components, which is an even greater task.
- Different structures of knowledge represent different types of intellectual tasks.
- Acquisition of instrumental and structural components of knowledge, which is a long and cumulative process.
- The process of adoption of knowledge is an organised institutional process and not a personal adventure.
- The process takes place in a situation of co-operative learning in which the adult has a supportive role, and the child has an active role.

Three types of mental development are important for school learning:

- the child's cognitive development;

- the development and inner organisation of knowledge;
- the cognitive development of the child through school learning.

Piagetian theory and the research inspired by it offer a masterful description of the child's cognitive development. Vygotsky (1956) also described this process of concept development. Developmental cognitive psychology (Mussen, 1983; Snow and Yalow, 1988) has also contributed to understanding the child's ability to process information.

The development and internal organisation of knowledge in different domains have their specific inner logic, which is the subject of the epistemology of each field and of the history of knowledge.

Vygotsky's theory is original in that it conceptualises, in addition, the cognitive development of the child through school learning. The theory postulates that at each stage of cognitive development (first type of knowledge) the child confronts the intellectual requirements contained in the development of the system of knowledge (second type of development). This confrontation produces a new, original type of development which cannot be reduced to the laws of cognitive development nor to the laws of the construction of scientific knowledge.

Under certain conditions, and with all the characteristics of cognitive behaviour that he possesses, the child assimilates structured knowledge in an original manner at each developmental stage. In the process, his system of thinking is changed. This is a lengthy, productive process that cannot be reduced to mere learning as memorising, but is at every moment "complex and truly an act of thinking in which the child's thought in its internal development is raised to a higher level of development" (Vygotsky, 1956, p. 216). Thus, just as the children expand their cognitive and communicative abilities through the adoption of language, they develop their thinking abilities through the adoption of the system of knowledge. Thus, the system of knowledge is another cultural medium that amplifies individual abilities because it provides the individual with a powerful intellectual tool. In relation to this form of development, Vygotsky states that "education can be defined as artificial development" (quoted in Schneuwly and Bronckart, 1985, p. 45).

The same ideas are also found in modern cognitive instructional psychology. Representatives of this new discipline, which actualises the theoretical perspective that considers schooling as a form of development modeled by culture, have this to say:

> "In contrast, more recent work on problem solving, done in a knowledge-rich domain, shows strong interactions between structure of knowledge and cognitive processes." (Glaser, 1985, p. 615)
>
> "Education is primarily an aptitude development programme. Intelligence is both a primary aptitude for learning in education and a primary product of learning in education." (Snow and Yalow, 1988, p. 559)

The relationship between mental development and school learning and the question of the nature of cognitive development through school learning are common to Vygotsky's theory and to modern cognitive psychology. A number of implications for the assessment of educational processes and outcomes can be extracted from this theoretical standpoint.

First, it is possible to develop a fruitful, critical analysis of the dominant models of assessment of educational outcomes. The structure of knowledge has not been assessed, nor has the process of concept formation, and there are no indicators of the instrumental and structural aspects of knowledge. The same is true for the technique of assessment.

Classroom problem-solving tasks are very rare, as are tasks requiring the use and transfer of conceptual knowledge. Finally, concerning the general doctrine of assessment, questions can be asked concerning the objectives of assessment, for example, the usefulness of static assessment of dynamic phenomena of school learning.

Second, these theoretical views make it possible to improve the interpretation of some of the existing indicators of educational outcomes. (If students do not acquire structured knowledge, it is probably because structured knowledge is missing from curricula, school textbooks, and educational practice. It may be possible to make similar assumptions about the limited ability of students to transfer acquired knowledge and the low level of validity of school knowledge assessed through standardized tests. In general terms, understanding the nature of the process of intellectual development through school learning would make it possible to reformulate the relationship between what is studied and what is assessed. It would also become possible to reformulate the so-called "curriculum/test overlap problem", the problem of the excessive similarity between curriculum content and achievement tests.

Next, the ground shared by Vygotsky's theory and cognitive instructional psychology presents the conceptual framework for the creation of complementary or alternative models of assessment. In terms of content, the measurement of educational outcomes necessarily implies the assessment of some components of the educational process. It would be necessary to develop indicators of the degree of structuredness (systemicity) of knowledge in curricula, school textbooks and educational practice. Indicators of co-operative learning in actual classroom practice could also be developed.

This approach opens up a broad spectrum of possibilities for enriching the assessment of educational outcomes. One might build instrumental and procedural knowledge into assessment batteries; one might introduce indicators of structuredness (integration) of conceptual knowledge; one might carry out a longitudinal study of the formation of the conceptual system. Thus, the content of assessment should not be limited to the curriculum of each grade level, but should also include the cumulative development of structure in one field of knowledge through several grades.

Basically, these ideas about innovations in the assessment of educational outcomes indicate that efforts should be directed towards the development of a new class of instruments for assessment. Tests of cognitive development through school learning (tests of "artificial development", in Vygotsky's terminology) should be elaborated. They should supplement existing psychometric tests, Piagetian diagnostics of the development of operations, experimental techniques for the study of problem-solving, and achievement tests. They should include tasks such as classroom problem-solving, a simulation of real-world problem-solving in certain domains of knowledge, transfer of knowledge, and tasks of integrated problem-solving in significant knowledge domains.

Clearly, the theories that conceptualise intellectual development through school learning highlight the need for a critical review of the existing models of assessment. In the context of lengthy, cumulative and qualitative processes of developing knowledge through schooling, basic characteristics of the existing systems – the linear development of school knowledge, Gauss's normal curve of score distributions, and the dominant static and`summative model – are questioned. Instead, new possibilities for dynamic assessment appear, including the assessment of qualitative degrees in the development of conceptual knowledge and of learning potential.

Finally, these ideas about models of assessment also pose the problem of a critical analysis of the general objectives of assessment. Instead of focusing on individual and group differences in achievement, new paradigms of assessment could be developed. They would have a feedback function that would allow for correction of the entire process of teaching, instruction and learning, and for undertaking special instructional efforts for groups and individuals. A second general goal would be the improvement of the ecological validity of assessment results.

There is, of course, the obvious difficulty of realising these ideas. Initially, they could be realised experimentally. Models of assessment could be developed that would at first complement existing ones, and that would serve specific groups of students and specific goals.

Metacognitive abilities

In dealing with cognitive development and learning, one inevitably comes up against the problem of metacognitive abilities. The analysis of the efficiency of school learning, attempts at computer simulation of cognitive behaviour, the testing of intelligence and experimental laboratory analyses of the process of problem-solving, all show that the efficiency of cognitive behaviour and the variance in cognitive achievement cannot be explained without postulating some ability of a higher order with a regulating, controlling and monitoring function. All this bespeaks the truth of the statement, attributed to A. Binet, that "it is not sufficient to possess intelligence, one should use it intelligently".

In the contemporary literature, metacognition truly appears as "a monster of obscure parentage and a many-headed monster at that" (Brown *et al.*, 1983, p. 124). The heterogeneity of the concept is evident in a simple enumeration of notions it encompasses:

- metacognitive subjective experience;
- all forms of self-regulating behaviour (regardless of the developmental level and the mechanisms of self-regulation);
- the knowledge of oneself as a knowing being;
- the executive and controlling aspects of information processing;
- self-reflection and self-consciousness;
- knowledge about one's own cognition.

The arguments below refer to metacognition in a restricted sense, adhering to two mutually related criteria: the specificity of Vygotsky's understanding of metacognition and the relevance of knowledge about metacognition for school learning and its effects. This makes it possible to focus on cognition in the process of systematic school learning.

Vygotsky clearly saw the beneficial influence of the general socio-cultural context on the development of metacognition. Specifically, the institution of schooling, in which an intellectual activity such as school learning appears as a specialised activity free from pragmatic constraints, is favourable to the development of self-reflection, self-awareness and self-knowledge. Besides, the general formula of individual development from inter-individual to intra-individual (Vygotsky, 1978, 1982-84; Ivic, 1987) is particularly applicable to the development of controlling and regulating functions. The planning, monitoring, control and evaluation of actions and psychological processes first take place in social interaction, where these functions are taken on by the adult. Metacognitive abilities appear in the course of the long developmental process of transition from the inter-psychological to the intra-psychological.

The second important aspect of Vygotsky's theory in this respect is his view that the most important parameter of development is not the development of each individual function but the change of relationships among individual mental functions, the integration of functions that creates new functional wholes. It is in this context that Vygotsky introduces the key concepts for his understanding of metacognition. These are the Russian concepts of *ossoznanie* (in English, "the grasp of consciousness") and *ovladanie* (in English, "control", "voluntary control", "deliberate mastery"). Vygotsky emphasizes that these two functions are the fundamental achievements of psychological development in the school years. He relates the origin of these functions directly to the development of the system of concepts in the process of school learning. In the course of ontogenetic development, the intellectualisation of all mental functions (action, memory, attention, etc.) takes place. In the process of adopting structured knowledge, intellectual operations become the object of intellectual analysis: "The grasp of consciousness (*prise de conscience; ossoznanie*) is an act of cognition in which the object of cognition becomes cognitive ability itself" (Vygotsky, 1956, p. 246). This is the essence of metacognition in the aforementioned sense of knowledge about cognition. In turn, this process makes *ovladanie* (*i.e.* control, deliberate mastery, monitoring, evaluation of one's own cognitive activities) possible. One can summarise these ideas as follows:

- The source of metacognition is in inter-individual control.
- Vygotsky conceptualises only the aspect of metacognition that can be labelled "knowledge about cognition".
- He deals with the mature forms of the metacognitive ability that are generated in the school years.
- This form of metacognition is an integral part of general cognitive development and appears in the process of establishing relationships between different mental functions. It basically consists in the intellectualisation of all these functions, including the intellect itself.
- This form of metacognition is necessarily generated in the development of the conceptual system.
- The manner of adopting this conceptual system favours the "grasp of consciousness" and the control of one's own cognitive processes. This occurs within the framework of a relatively autonomous activity of school learning, in interaction with socially shaped systems of knowledge rather than in direct contact with reality, and in the process of verbal learning which takes place in a co-operative learning situation.

Because of the relatively immature state of scientific knowledge about metacognition, it is difficult to deduce conclusions useful for the assessment of educational processes and outcomes. It can only be said that the metacognitive aspects of educational outcomes cannot be ignored. In future years, theoretical and empirical research on metacognition will be of primary significance. In addition to experimental investigations, research on the nature and function of metacognitive abilities in school learning would be valuable. Vygotsky's specific conceptualisation of metacognitive processes, which relates the appearance of their mature forms to the period of schooling and to school learning, can also serve to focus research on metacognitive ability in the school environment. This may be helpful for defining objective indicators of the presence or absence of metacognition in the organisation of school learning.

Piaget and indicators of educational outcomes

Whereas Vygotsky placed the axis of intellectual development in the process of systematic instruction in school, Piaget accorded little significance to it. Nevertheless, the problem of assessing educational processes and outcomes can be discussed from the perspective of Piaget's theory, which is, in essence, a theory of the psychogenesis of knowledge. He defines psychogenesis as the development of cognitive structures and of cognitive abilities, and extrapolations from this theory can be fruitful because it is a masterly description of the active role played by the child in developing his own cognitive abilities and in acquiring knowledge.

With respect to the development of knowledge, one might say that Vygotsky, through his analysis of the system of knowledge which the child adopts at school, has elucidated the notion of "the object of knowledge", while Piaget can be credited with elucidating the notion of "the subject of knowledge". The following central elements of Piaget's theory are relevant to the process of school learning and its outcomes:

- the study of the development of intellectual operations as the basic tools in the acquisition of knowledge;
- the description of the genesis and development of these operations;
- the analysis of the epistemological nature of many forms of knowledge;
- the description of the development of concepts in many scientific disciplines.

The components of Piaget's work that illuminate the acquisition of school knowledge and the assessment of the outcomes of this process are: the formation of the system of intellectual operations; the relationship between cognitive structures and content of knowledge; the relationships between achievement, understanding and metacognition.

Formation of the system of intellectual operations

The formation of the system of intellectual operations lies at the heart of Piaget's theory. The corpus of knowledge generated by this theory has resulted in a masterly description of the internal logic of ontogenetic development of cognitive abilities. Hence, attempts to create applied instructional programmes, or even to assess outcomes of programmes, must take Piaget's theory into consideration. The reasons for this are as follows.

In each developmental period, the system of intellectual operations furnishes a basic cognitive tool for acquiring knowledge. The dynamics of the acquisition of knowledge and the effects of its adoption will depend on the degree to which intellectual operations are developed in successive periods. Overall, the development of intellectual operations leads towards the formation of a system of intellectual operations. Even if Piaget's theory about the general stages of child development is questioned, the idea that the formation of the system of intellectual operations is a process of qualitative transformations, rather than of linear quantitative growth, is generally considered valid.

For this reason, special attention should be given to certain revisions of Piaget's stage theory that postulate separate "within-domain stages of development" (Gelman and Baillargeon, 1983) or conceive development as a process "characterised by structural change" and "paradigm shifts" (Bruner, 1985, p. 600). In any case, as Bruner says, the process cannot be reduced to "the simple accretion of information".

Piaget's theory conceptualises the *macrogenesis* of intellectual operations, *i.e.* their appearance and development in the course of ontogenesis. In this respect, Piaget's investigations differ significantly from most research on intellectual processes in cognitive psychology, in developmental cognitive psychology, and even in a good deal of experimental instructional psychology, which is usually restricted to the study of *microgenesis*, *i.e.* the processes implied in the solution of each particular intellectual task. This is what makes this aspect of Piaget's theory relevant for school learning, which is also a macrogenetic process.

The development of the repertoire of intellectual operations (logico-mathematical operations, spatial-temporal operations, propositional operations) has been studied in the framework of Piaget's theoretical orientation, and this work must be taken into account in curriculum development and in constructing tasks for the assessment of educational outcomes. Piaget's understanding of the factors of development and of the development of intellectual operations is of great significance for the study of instructional processes and its outcomes. The child enters every new encounter with knowledge equipped with an existing repertoire of operations, and new knowledge is acquired through the active construction of this knowledge with the aid of cognitive tools at the child's disposal.

A series of consequences follows from this most fundamental of Piaget's theoretical assumptions:

- The child always has inner motivation for acquiring knowledge.
- The developing child selects knowledge to which he is sensitive in a certain period of development.
- New knowledge is not received as it is, but is assimilated in a specific manner.
- New knowledge is gained not only as a result of outside influences but also through the inner balance and co-ordination of the existing cognitive schemata and knowledge. Thus, advances in the development of knowledge are possible even when outside influences stop, because internal processes of integration and co-ordination take place in the cognitive system of the child.
- The result of every process of acquiring new knowledge is not simply the copying of knowledge that exists in a discipline or a school subject, but is the original result of the interaction of the developing child's cognitive schemata with this discipline-bound knowledge. These factors and processes of cognitive development are essential for understanding instructional processes and their outcomes.

Piaget was not concerned with the study of the processes of cognitive development through school learning, probably because he underestimated the significance of this process. However, Piaget was often concerned with the inner logic of the constitution of particular domains of knowledge (see, for example, Piaget, 1967).

The inner logic of the development of the child's thinking is at the centre of Piaget's theoretical interests. In this context, school learning cannot be a process of simple accumulation of knowledge. It necessarily entails thinking, *i.e.* the active reconstruction of thought, or, as Piaget wrote in his study of educational problems: "*to understand just* *means to invent or to make the reconstruction by the process of re-invention.*" (1972, p. 24, italicised in the original)

From this brief review of Piaget's theory of development – the formation of the system of intellectual operations – it is possible to extract useful implications for assessing educational outcomes:

- The theory presents a basis for a critical analysis of the dominant models of assessment of educational outcomes. It would criticise the dominant types of tasks in achievement tests, which are usually reproductive and atomised and which test knowledge not integrated in the child's cognitive schemata. Moreover, it would question the linear-additive model of acquiring knowledge, which is the underlying principle in the construction of achievement tests.
- For assessing outcomes, the theory justifies the construction of new types of tasks that engage the child's thinking. These might include tasks of understanding, concept comparison, interconnection of knowledge, invention of new examples to illustrate acquired concepts, critical analysis of different statements, and application of the same intellectual operations to different conceptual content. The tests would have to offer the student the opportunity to show himself as an active and independent learner.
- The conceptualisation of cognitive development as a series of qualitative transformations in the system of intellectual operations has many implications. One is that a new model, opposed to the dominant linear-additive models for the assessment of outcomes, is needed. In such a model:
    - The development of knowledge would be tested as macrogenesis, which transcends specific school subjects.
    - This would make it possible to address qualitative transformations of knowledge, *e.g.* higher levels of knowledge, ways of integrating knowledge, higher types of definitions and concepts, and new cognitive operations.
    - These new aspects could be tested through a new type of outcome measure and not only by a quantitative score.

Cognitive structures and the content of knowledge

The relationship between cognitive structures and the content of knowledge is an important aspect of Piaget's theoretical system. It is also important for conceiving instructional programmes. While the core of school curricula is composed of the content of various bodies of knowledge, the cognitive structures of the students cannot be ignored. Piaget gave primacy to the application of cognitive structures to solve problems and understand different contents. He found identical cognitive structures (such as the formation of logical classes on the basis of some common characteristic) in children of different ages, although the concrete content varied (classifying pictures of familiar animals or abstract geometric shapes). In order to explain this phenomenon, Piaget introduced the concept of (horizontal and vertical) *décalage*. Horizontal *décalage* is the time lag in applying the same structure to different contents. Here, what is specific in the content of knowledge is perceived as a hindrance, a barrier to applying the cognitive structure. This is theoretically inelegant, and, when applied to school learning, it ignores the specificity of the content of knowledge.

A revision of the relationship between structure and content is offered by Montangero (1985): "Rather than considering the problems of horizontal decalage in negative terms of 'obstacles', 'resistance', or 'friction' preventing the subject's coherence to exert itself, I suggest to address the question in positive terms of construction of new and more complex inferences" (p. 39). Montangero's analysis and his experimental demonstration using the example of the concept of time, opens up new possibilities for posing and solving the problem of relationships between cognitive structures in the developing child and the various contents of school curricula.

This revision of Piaget's theory becomes even more interesting for instructional theory when considered in conjunction with the revision of the theory of general stages of development. The conceptual innovations that consist in introducing the strategies and procedures of intellectual problem-solving into the repertoire of cognitive behaviours should also be taken into consideration (Inhelder and Piaget, 1976; Inhelder and Piaget, 1979; Cellerier, 1979; Montangero, 1985). Montangero clearly stated the significance of these innovations for the process of school learning in different domains of knowledge: "Logical structures are thus not constructed independently from the different domains which they can organise. Moreover, when the operatory systems are constituted, they enrich the understanding of this domain" (1985, p. 60). These new theoretical solutions are directly relevant to the notion of instructional programmes that ensure two-way interaction between the existing cognitive structures of the child and new conceptual contents. In this interaction, the existing structures are stabilised, generalised, and, therefore, developed. This is significant for the assessment of student achievement.

The concepts of procedure and strategy in problem-solving also open up the possibility of their exploitation for school learning. If mental operations are the basic instruments for problem-solving and for understanding different domains of knowledge, then procedures (concrete techniques of solving given problems) and strategies (systematisation of procedures and sequences of concrete procedures) are more flexible and adaptable. These cognitive tools depend to a greater extent on specific domains of knowledge. The crucial issue for understanding the process of school learning and its influence on cognitive development is the interaction of the more concrete forms of cognitive behaviour, which include strategies and procedures of problem-solving, and more general behaviours such as operations. It could be said that the student enters new learning (*i.e.* he acquires new contents) through cognitive operations. It could also be said that concrete strategies, procedures and cognitive behaviours act on the generalisation of mental operations, co-ordinating and interconnecting them into more complex systems.

The revised Piagetian theory offers an instrument for the fruitful analysis of the process of school learning and its effects, as is shown by the numerous empirical investigations of the operations, strategies and procedures used by children in different developmental periods when solving various intellectual tasks. These analyses describe, precisely and in detail, the epistemological nature of particular contents of knowledge, *i.e.* their structural nucleus. The investigation of different types of conservation tasks and their relationships carried out by Inhelder *et al.* (1974) offer an excellent example. This work is relevant for curricula and for the typology of tasks included in the batteries of tests used to assess educational outcomes.

It is possible to extract implications for student assessment from the analysis of the relationship between cognitive structures and the content of knowledge:

- One can identify the epistemological nature of much of the conceptual content in structural analyses of intellectual tasks and in the analyses of operations, strategies and procedures. This makes it possible to enrich both curricula and types of tasks in assessments by including types of problems whose structural characteristics, developmental levels and intellectual demands are known.
- Knowledge about the epistemological nature of developmental tasks makes it possible to pose the problem of the transfer of learning in a new way, both for the process of learning and for the method of assessment. Transfer is possible only where there are structural similarities between intellectual tasks and where the contents based on the same cognitive structure are sufficiently varied in the learning process.
- Horizontal *décalage* can be positively interpreted as the process of making new and more complex inferences through the variation of different contents based on isomorphic or similar structures. Interpreted in this way, horizontal *décalage* allows a developmentally fine grading of the acquisition of operations and strategies which serve for solving structural intellectual problems.

Achievement, understanding and metacognition

The relationship between success (or achievement) and understanding is a theoretical component of some of Piaget's later works (1974*a*; 1974*b*). It is important for understanding the nature of educational outcomes and for defining their different levels. According to Piaget's theory, understanding is achieved through the process of assimilation, the integration of new elements into existing cognitive schemata. As a result, the key element in Piagetian models of assessment is the emphasis placed on understanding the problem (revealed by an analysis of the argument given in the answer), rather than on mere success. In classic psychometric achievement tests, instead, success or failure are scored on the basis of individual tasks.

Piaget's later work addresses the relationship of practical knowledge (*savoir-faire*) and conceptual knowledge. Another way to formulate this opposition is as technology-science versus action-thought. All empirical research about the ontogenetic development of the relationship between practical achievement and "conceptual understanding" is also relevant. These distinctions are useful for understanding educational processes and outcomes:

- Two kinds of achievement are realised in the process of learning: "practical" or "procedural" knowledge and conceptual knowledge.
- Conceptual knowledge comes developmentally later and is based on practical achievements.
- Conceptual knowledge, when it appears, has a "backwash" effect on practical knowledge.
- These two kinds of knowledge have different objectives: practical success and understanding.

The most complex stage of developing conceptual knowledge is the "grasp of consciousness" (*prise de conscience*). When "the subject becomes able to structure reality operationally, he remains for a long time unaware of his own cognitive structures: even when he uses them for usual purposes or even when he attributes them to objects and events so as to explain them causally, these structures do not become topics of his reflection (they are not thematised) until the highest level of abstraction is reached" (Piaget, 1974*a*, p. 234). This level of abstraction is the level of conceptualisation in the real sense, the level where conceptual systems are formed. At that point, at around age 11-12, consciousness begins controlling and programming actions, orienting them, anticipating them, and achieving an understanding of them.

In other words, Piaget is concerned with *metacognitive abilities*. Like Vygotsky, Piaget relates metacognition to mature stages of development and to the appearance of conceptual systems of knowledge. He also understands it as "knowledge of one's own cognition". He describes the functions of these abilities in similar terms. The qualitatively higher level is reflected by the fact that, when discussing the regulation of cognitive behaviour, Piaget relates intentional and conscious regulation only to the third and highest level (above autonomous and active regulation). These distinctions are useful for clarifying the concept of metacognition. The third and highest level plays a role in school learning, but this has yet to be carefully studied.

Piaget's theoretical distinction between two types of acquisitions – success (achievement) and understanding – also has major implications for the assessment of educational outcomes:

- Tests for assessing educational outcomes contain and should contain two categories of achievement: practical knowledge (elsewhere, *savoir-faire*, procedural knowledge, etc.), which is characterised by practical success; and conceptual knowledge, which is characterised by understanding. What is new from a theoretical perspective is the knowledge of the nature, range and limitations of each of these categories of knowledge. This makes it possible to formulate more clearly the purpose of including each type of achievement in assessment tests. The tasks used to test practical success and conceptual understanding can also be developmentally graded according to degree of achievement.
- Tasks that test understanding are the key to defining the qualitative levels of the development of knowledge. They should dominate in the assessment of the development of structural knowledge during the entire period of schooling. Experimental studies of the process of understanding and the explanations given by subjects are a rich source of ideas for the construction of tasks that test this process. This process covers all the levels of conceptual understanding that lead to metacognitive knowledge (the naming of reasons, causal explanations, the explanation of modes of production of different phenomena, the understanding of cause and effect, etc.).

Tentative conclusions

This chapter began with the assumption that theories of mental development can be relevant for the critical examination of approaches to the assessment of educational outcomes and for the development of complementary and/or alternative systems of assessment. With this goal in mind, several components of Vygotsky's and Piaget's theories of mental development were reviewed, along with aspects of modern cognitive psychology and instructional theory.

Certain tentative conclusions about assessing educational outcomes were drawn from these theories by extrapolating from theoretical and empirical knowledge about cognitive development. The rational basis and the justification for these extrapolations derive from the fact that both the theories of mental development and the theories of school teaching and learning are concerned with the development of knowledge. In the first case, the issue is addressed from the perspective of the development of cognitive abilities and, in the second, from the perspective of the acquisition of a system of knowledge. In instructional psychology, the centre of attention is the attempt to synthesise these two perspectives. School learning is viewed as a contemporary medium for realising cognitive development. This was another of Vygotsky's innovative ideas.

The extrapolations mentioned above can be summarised as follows:

- School learning is conceived as the formation of systems of knowledge, and the basic assumptions of the system of assessment must reflect this concept.
- The formation of systems of knowledge occurs through a process of qualitative transformations (which the system of assessment must also reflect).
- The basic goal of assessment is to offer a basis for corrections in the educational process.
- The overall goal is the improvement of the ecological validity of the system of assessment of educational outcomes.

Critical analysis of the existing systems of assessment

Knowledge gained from the theories of mental development can serve as a useful starting point for a critical analysis of the existing dominant models of assessment:

- in terms of the content of assessment (*e.g.* the dominance of tasks requiring the reproduction of knowledge, and the corresponding lack or small number of tasks that test complex kinds of knowledge such as understanding, integrated knowledge, transfer, classroom problem-solving, procedural knowledge);
- in terms of methods and techniques of assessment (*e.g.* static and summative models of assessment rather than dynamic ones; multiple choice answers and paper-and-pencil tasks, which reduce the repertoire of possible tasks; assessment situations which deprive students of the equipment which they use in real-world problem-solving situations);
- in terms of the implicit assumptions on which these models of assessment are based (*e.g.* viewing the process of formation of knowledge as a linear-additive process, using Gauss's statistical model of the normal distribution of results as the basic referential framework for the interpretation of results);
- in terms of objectives of assessment (*e.g.* concern with the establishment of individual differences, most often for purposes of selection; with the prognosis of school learning, most commonly using achievement tests of a similar type in later stages of schooling), which rarely deal with the goals of analytic feedback that shed light on the process of school learning, the improvement of ecological validity, etc.

Interpretation of assessed educational outcomes

Theoretical and empirical knowledge about cognitive development can aid in understanding the findings and interpreting the indicators obtained by using the existing systems of assessment. They make it clear that the indicators called process indicators in the INES Project have to be related to indicators of outcomes, because this is the only way to look at the mode of producing outcomes in the educational process. In other words, a connection must be established between what has been taught and what is assessed.

This holds true for negative as well as for positive correlations. In the case of the former, if it is established in the assessment of outcomes that acquired knowledge is partial and not integrated into systems, it may be that neither the curricula nor the process of instruction draw on knowledge systems. In the case of the latter, heightened results on an achievement test may be the simple result of the fact that the classroom learning is too test-oriented, so that it is almost entirely reduced to preparation for achievement testing.

Conceptual bases for new models of assessment

It follows that the theories of mental development, the knowledge of basic psychological research on cognitive development, and the practice of assessing cognitive development can serve as a useful conceptual starting point for opening up new perspectives for outcomes assessment. However, these new perspectives are not possible without a different understanding of school learning. The instructional process is the process of interaction between the cognitive abilities of a developing child and a system of knowledge (and the intelligence built into that knowledge). The ideal result of this lengthy process is the formation of expert competence consisting of structured systems of knowledge, modes of thinking used in every major domain of knowledge, a system of procedural knowledge, and the ability to regulate cognitive activities (*i.e.* expert levels of metacognition).

In the process of school learning, cognitive development consists of a series of changes that take a novice through different stages of expertise. The basic function of assessment would be to define the stages in the development of expertise. The following paragraphs briefly offer some ideas on how a new system of assessment based on the theories of mental development might be formed.

New domains of assessment

Traditional systems of assessment assess knowledge acquired in particular school subjects and ignore significant aspects of educational outcomes. These include: aspects of procedural knowledge (whether related to school subjects or more general); skills in using equipment for intellectual work (apparatus, information technology); skills in using various kinds of information sources (libraries, information centres and data banks, secondary and tertiary sources of information); the ability to create personal documentation and organised sources of information; metacognitive abilities; and ability to do team work.

New content in test batteries for assessment

The theoretical work discussed above has resulted in greatly increased understanding of the nature of knowledge. Much is now known about the structural characteristics of many forms of learning, about the intellectual demands posed by the acquisition of knowledge, and about the developmental levels involved. Such precise epistemological knowledge is of great value for composing curricula and for determining the order of learning as well as for developing tasks for assessing acquired knowledge. On this basis, the repertoire of tasks included in diagnostic instruments for assessing educational outcomes can be enriched by creating: tasks that test conceptual understanding (giving the reason for a judgement, explaining the way a phenomenon is produced, understanding the process that leads from cause to effect); tasks that test the integration of knowledge (understanding whether concepts are logically subordinate or logically superior, tasks of logical calculus, tasks requiring placing specific cases in general categories or creating new examples of a conceptual category); transfer tasks (transferring acquired knowledge to new cases, applying acquired knowledge to real-world situations); tasks of instrumental knowledge (acquaintance with the methods and techniques in a domain of knowledge); tasks of complex procedural knowledge and its application in new problem situations.

Such tasks already exist sporadically in many achievement tests. However, by systematically constructing this type of assessment tasks, a new category of instruments for assessment could be developed. Specifically, an instrument could be created to assess thinking (problem-solving skills, high-level thinking skills, higher order learning skills), which could in turn be applied to some of the large domains of knowledge (*e.g.* experimental thinking, axiomatic thinking, historical thinking). This new instrument could supplement the classical psychometric tests of intelligence, the operational tasks used in Piaget's diagnostics, and the achievement tests which dominate in the assessment of school learning. The new instrument would fall between intelligence tests and achievement tests. The variable tested using this instrument would be the form of thinking that is developed by adopting the structured body of human knowledge.

New methods of assessment

The methods of assessment are closely related to the content of assessment. It would therefore be necessary, in elaborating new paradigms of assessment, to change methods as well. Some ideas for change can be derived from the preceding discussion. Instead of depriving subjects of equipment for intellectual work, the necessary instruments and information sources (or a standard set of such tools) could be placed at the candidate's disposal. Primacy would then be given not to the reproduction of knowledge, but to problem-solving, the use of auxiliary tools, transfer and the application of knowledge. In addition, tasks could be created, on the basis of knowledge about the epistemological nature of various kinds of knowledge and about the order of acquisition of this knowledge, that would take account of a progression in cognitive developmental.

New models of assessment

At another level, the principles guiding assessment could also be revised. Instead of the linear-additive model of knowledge acquisition, a qualitative model of knowledge could be developed, which would make it possible to study the development of knowledge using the criterion of qualitative change. Thus, instead of the reproduction of knowledge, the specific developmentally graded forms of classroom problem-solving and real-world problem-solving would be tested.

To replace the statistical model based on the normal distribution of scores on standardized student achievement tests, a model based on learning potential could be developed. It would be a dynamic, rather than a static model of assessment. Theoretically, such a model would be based on Vygotsky's concept of "zone of proximal development".

Changes in the objectives of assessment

The dominant objectives of present-day tests relate to determining individual and group differences in achievement. New assessment paradigms could be based on alternative objectives: obtaining precise information on the instructional process, which could be used to perfect this process; obtaining information about the qualitative level in the development of knowledge, which would be used to create programmes for further learning; obtaining information about the learning potential of the individual, so as to create a programme for further learning adapted to that potential and the learning already achieved. New assessment models would also be used to attain goals that are complementary to those that are presently dominant, such as selecting specific populations of students (*e.g.* for experimental learning programmes, for advanced courses, for slow learners, or for certain international instructional programmes). In some cases, even personalised assessment can be justified.

Complementary and/or alternative ideologies of assessment

The implicit ideology of the currently dominant, psychometric models of assessment and achievement testing stems from the socio-cultural context in which these models originated. Basically, this ideology confirms existing individual and group differences, with the goal of "neutral" prognosis of future achievement and, eventually, student selection for educational and occupational careers. In accordance with certain new ideas in education (*e.g.* employment of all human resources, quality education for all), new presuppositions of assessment can be developed, such as assessment with the objective of having all students reach optimum achievement or of ensuring all segments of the population a certain minimal level of competence.

Bibliography

BROWN, A.L., BRAMSFORD, J.D., FERRARA, R.A. and CAMPIONE, J.C. (1983), "Learning, Remembering and Understanding", in Flavell, J.H. and Markman, E.M. (eds.), *Handbook of Child Psychology*, Vol. 3: *Cognitive Development*, pp. 77-166, Wiley, New York.

BRUNER, J.S. (1985), "Preface", in Chipman, S., Segal, J. and Glaser, R. (eds.), *Thinking and Learning Skills*, Vol. 2: *Research and Open Questions*, Laurence Erlbaum and Associates, Hillsdale, NJ.

CELLERIER, G. (1979), "Structures cognitives et schèmes d'action", *Archives de Psychologie*, Vol. 47 (180), pp. 87-106.

ERAUT, M. (ed.) (1989), *International Handbook of Educational Technology*, Pergamon Press, Oxford.

GELMAN, R. and BAILLARGEON, R. (1983), "A Review of Some Piagetian Concepts", in Mussen, P.H. (ed.), *Handbook of Child Psychology*, Vol. 1, pp. 167-230, Wiley, New York.

GLASER, R. (1985), "Thinking and Learning Skills", in Chipman, S., Segal, J. and Glaser, R. (eds.), *Thinking and Learning Skills*, Vol. 2: *Research and Open Questions*, Lawrence Erlbaum and Associates, Hillsdale, NJ.

INHELDER, B. and PIAGET, J. (1976), "Des structures cognitives aux procédures de découverte", *Archives de Psychologie*, Vol. 44 (171), pp. 57-72.

INHELDER, B. and PIAGET, J. (1979), "Procédures et structures", *Archives de Psychologie*, Vol. 47 (181), pp. 165-176.

INHELDER, B., SINCLAIR, H. and BOVET, M. (1974), *Apprentissage et structures de la connaissance*, Presses universitaires de France, Paris.

IVIC, I. (1987), "Le social au cœur de l'individuel", in CRESAS, *On n'apprend pas tout seul: interactions sociales et construction des savoirs*, European Science Foundation, Paris.

IVIC, I. (1989), "Profiles of Educators: Lev Vygotsky", *Prospects*, Vol. 19 (3), UNESCO, Paris.

MONTANGERO, J. (1985), *Genetic Epistemology: Yesterday and Tomorrow*, City University of New York, New York.

MUSSEN, P.H. (ed.) (1983), "Cognitive Development", in Mussen, P.H. (ed.), *Handbook of Child Psychology*, Vol. 1, Wiley, New York.

OECD (1992), *High-Quality Education and Training for All*, OECD, Paris.

PIAGET, J. (ed.) (1967), "Logique et connaissance scientifique", *Encyclopédie de la Pléiade*, Gallimard, Paris.

PIAGET, J. (1972), *Où va l'éducation?*, Denoël-Gonthier, Paris.

PIAGET, J. (1974*a*), *Réussir et comprendre*, Presses universitaires de France, Paris.

PIAGET, J. (1974*b*), *La prise de conscience*, Presses universitaires de France, Paris.

SCHNEUWLY, B. and BRONCKART, J.P. (eds.) (1985), *Vygotsky aujourd'hui: textes de base*, Delachaux et Nestlé, Neuchatel.

SNOW, R.E. and YALOW, E. (1988), "Education and Intelligence", in Sternberg, R.J. (ed.), *Handbook of Human Intelligence*, pp. 493-587, Harvard University Press, Cambridge, MA.

VYGOTSKY, L.S. (1956), *Selected Psychological Works* (in Russian), APN RSFSR, Moscow.

VYGOTSKY, L.S. (1978), *Mind and Society*, Harvard University Press, Cambridge, MA.

VYGOTSKY, L.S. (1982-1984), *Collected Works*, 6 vols. (in Russian), Pedagogika, Moscow.

Chapter 11

The Implications of Cognitive Psychology for Measuring Student Achievement

by

Herbert J. Walberg
University of Illinois, United States

and

Geneva D. Haertel
Palo Alto, California, United States

Educators find themselves confronting rapid changes in the types of knowledge and achievement of the global community, as well as their particular society, value. Models of educational accountability are being developed that attend carefully to measuring students' higher level thought processes in specific content domains. Reforms in school structure, curriculum, instruction and assessment have become the topic of many educational conferences, books, and scholarly papers (Finn, 1991).

Educators and the business community have joined forces in expressing concern that students be able to function effectively in an increasingly technical society, emphasizing the need for higher level thought processes in all school subjects. Policy-makers have placed a premium on obtaining accountability information to assess individual student and group performance at the classroom, district, state and national levels. The decisions based on assessment data have "high stakes" consequences for students, their teachers, administrators and the economic future of their countries. Educators have sought information and direction on the use of assessment for monitoring student growth and group performance in school subjects. Developments in cognitive psychology have supplied a new perspective on the assessment of abilities and achievement specific to schooling.

The purpose of this chapter is to describe the impact of cognitive psychology on the measurement of student achievement. It begins with a brief history of the influence of cognitive psychology on the measurement of intellectual abilities and aptitudes and presents the contribution of cognitive psychology to the measurement of abilities and aptitudes. It then describes the impact cognitive psychology has had on the measurement of student achievement.

Glaser (1984) stated that when measuring abilities and generalised achievement a substantive theoretical underpinning is useful, but as one moves towards achievement in specific content areas, the need for strong substantive theories of domain-specific achievement becomes critical. Cognitive psychologists have provided a strong argument for domain-specific achievement tests. They have identified paths of expertise ranging from novice to expert performance in a variety of subject areas. The terms "declarative" and "procedural" knowledge have become common parlance among educators. Automaticity, a critical feature of the rapidity with which cognitive functions are executed, has become a key component in understanding how knowledge is proceduralised. Novice versus expert performance, declarative and procedural knowledge and automaticity, are constructs that cognitive psychologists have developed and applied to the design of achievement tests.

The impact of cognitive psychology on the measurement of intellectual abilities: a selective history

Early psychologists made a systematic effort to catalogue human abilities. Sir Francis Galton developed statistical methods that allowed him to rank individuals according to their intellectual and physical capacities. In his view, sensory discriminations were the key to determining an individual's intelligence, and the most capable individuals would exhibit especially rapid and accurate sensory discriminations (Boring, 1929). Gardner (1983) noted that psychologists quickly broadened their definition of intellectual power to include more complex tasks involving linguistic and abstract abilities. These were captured in Binet and Simon's intelligence test (Binet and Simon, 1905). The Binet and Simon test was developed to identify children who were not able to perform well in formal school situations and might require placement in a special education class. Binet did not believe intelligence was a unitary trait; however, the fact that a single score was used to report the test results reified the construct. For many, intelligence was an unchanging, single trait that an individual possessed from birth to death.

Models of human ability

After the development of the Binet and Simon intelligence test, much of the research on human abilities was conducted by psychometricians. For nearly 75 years (approximately from 1905 until 1980), psychometricians argued over the true structure of mental abilities (Sternberg, 1990). Numerous models were put forth describing the structure of intellectual abilities, including those of Burt, Cattell, Guilford, Jensen, Spearman, Thomson, Thurstone and Vernon (Gustafsson, 1984). These models were based on multivariate statistical analysis, in particular factor analytic procedures. Some of the models claimed the presence of a general factor of intelligence (Spearman, 1927); others did not permit a general factor of intelligence to emerge, but allowed for the emergence of group factors (Thurstone, 1938), or specific factors (Vernon, 1971). There were also hierarchical models (Burt, 1940; Vernon, 1971), and models with multiple factors that were treated as being of equal generality (Guilford, 1967). These various models of intelligence were based primarily on the statistical method used to rotate the factors. The emphasis was on the mathematics applied, not on the psychological attributes (Sternberg, 1990).

Many of the tests that were administered in order to develop these models were very similar, but the psychometric methods were increasingly complicated. As Sternberg asserts, "The metrics advanced; the psychology did not" (1990, p. 204).

A taxonomy of human abilities

Snow (1986) cites Carroll's (1985) taxonomy of human abilities, which is hierarchic and includes the most heavily researched abilities. The validity of the ability factors has been confirmed using factor analytic data. The general or "g" factor is at the top of the hierarchy. There are three second order factors: crystallised intelligence, which is a generalised educational achievement factor; a fluid intelligence factor, which includes general deductive reasoning and logical reasoning; and a general visual perception factor which represents figural-spatial-imagery demands. Other abilities represented in Carroll's taxonomy include the production of general ideas or associative fluency, auditory perception, memory and speed. Carroll's taxonomy provides one of the more recent classifications of human abilities that have been documented through factor analytic research.

Research results have repeatedly confirmed the correlations between the major abilities and performance in school. One of the highest correlations is the relationship of crystallised ability with learning from instruction (Snow, 1986). Performance in any particular school subject requires a variety of abilities. For example, preparing a recipe in a home economics course may require visual perceptual discriminations as well as memory during the initial phases of problem representation. As the student proceeds through the recipe, other abilities, including reasoning skills, verbal comprehension and knowledge about cooking, may be necessary. Both fluid and crystallised abilities will be used as the student proceeds to make the recipe. Snow (1986) describes how different abilities are orchestrated in performing school tasks. When the school task relies heavily on a student's prior knowledge, then the crystallised abilities will be the critical attributes. When the school task is more novel, then the fluid abilities will be dominant. Visual-spatial abilities will be employed when necessary for problem solution, but have primarily been related to achievement in areas such as dentistry, architecture, and shop courses.

Multiple intelligences: a new perspective

Society's understanding of the construct of intelligence has evolved from that of a single unitary trait towards a recognition that there are multiple types of intelligence (Gardner, 1983). Gardner identified seven relatively independent types of intelligence or competencies, including: linguistic, musical, logical-mathematical, spatial, kinaesthetic, interpersonal and intrapersonal competencies; each of these is described as having a developmental trajectory. Individuals vary in the degree to which each type matures; therefore, individuals can reach different degrees of competency depending on the social context in which they live. The final states of competency reached are a function of the individual's genetic endowment interacting with the social context. Gardner's seven types of intelligence or competencies are not based on elaborate statistical models like those used by psychometricians; rather, he used findings from human development, neurology, and cross-cultural research. He examined populations of normal individuals, as well as brain-damaged individuals, idiots savants, prodigies and artists in order to identify multiple types of intelligence (Kornhaber *et al.*, 1990).

Csikszentmihalyi and Robinson (1986) assert that societies teach their children bodies of facts, theories, and skills from many domains of knowledge that are authentically valued. These domains range from how to jump rope to doing quadratic equations. Competencies, or types of intelligence, develop in the light of the skills, values, and knowledge the societies teach their children.

BEST COPY AVAILABLE

A developmental perspective

Context is a critically important influence on abilities. Some psychologists study abilities by watching how they develop and integrate with the passage of time (Piaget, 1963). Others have examined the way that environmental pressures have impacted the development of competencies and especially domain-specific expertise (Keating, 1990). Still others have studied the development of biologically constrained mechanisms and the effect of environment on them (Vernon, 1990). In all cases the developmental perspective recognises that cognitive processing is different at various stages of children's growth. These differences occur in sensory intake and short-term memory, long-term memory, and the interaction of system components (Michel, 1990).

For example, Thomas (1985) documented differences in the number of "chunks" of information that children of different ages can remember. At 18 months, a child can remember one chunk, by age five he or she can retain four, adults can typically recall seven. These developmental differences in short-term memory capacity represent changes in processing capacity with implications for the design and formatting of test items. Long-term memory changes with age, such that older children have more complex semantic networks and a greater number of associations among knowledge units. More rehearsal and mnemonic techniques are used as children age, and the rapidity with which information is processed also increases. Michel (1990) concludes that as children age the human information system increases in complexity, speed and integration. These conclusions suggest that tests need to be constructed that are sensitive to the developmental continua that underlie all abilities. Test items that do not recognise limitations in children's short-term memory, speed of processing and degree of complexity in long-term memory are not likely to produce valid results. A developmental trajectory is critical to the development of a model of human abilities, as well as to the design of assessment methods.

Models of information processing

During the 1970s, psychologists became interested in how individuals process information when they are completing complex, intellectually demanding tasks. The individuals who conducted this research were not psychometricians but would best be described as cognitive psychologists. Their work was directed at developing and testing new theories describing "the internal workings of the human cognitive system, the mental events and processes that connect stimulus and response in an item, test or task performance and thus to penetrate the black box of the EPM [educational psychometric measurement] models" (Snow and Lohman, 1989, p. 269). Laboratory tasks were developed that allowed researchers to distinguish selected cognitive processes and then follow an individual's performance stage by stage in terms of his perception, memory and attention. Using approaches such as these, psychologists developed cognitive information-processing models (CIP models). These models detail what happens during each component process and specify how the processes combine. The model is then operationalised and compared against rival models. The process is repeated and the CIP model is updated as needed. Sternberg (1985) and Dillon (1985) specify the types of CIP models that have been developed. Some of the kinds of processes that have been explored using CIP models include: word fluency, phonetic coding, episodic memory, and spatial relations. By identifying component processes, cognitive psychologists can locate where students have problems. It is possible to diagnose errors caused by lack of attention, perceptual problems and other processing deficiencies. Results from these studies have been applied to develop models of reading, mathematical and verbal abilities, as well as second-language learning. Information processing models became a major source of new evidence about human abilities.

The identification of information-processing models requires that test items be diverse, employing a variety of item formats so that speed of execution, as well as complexity of response might be recorded. It is important that tests be diagnostic and employ methods to discern different strategies to solve problems. Examinations that use a multiple-choice format, or true/false format as well as forced-choice methods may prove less useful for diagnosing deficiencies in processing abilities, whereas constructed response formats, foils and use of prompts may be more suitable.

Implications for the assessment of school achievement

Briefly traced, the history of human abilities research has been marked by a number of important trends. First, there was a shift from regarding intelligence as a fixed, unitary trait towards recognising that there are multiple intelligences. Second, it was understood that there is a developmental continuum underlying ability factors and that an individual's endpoint, in terms of competency, depends on the context in which he or she are immersed. Third, school tasks, such as reading comprehension, are the product of a myriad of componential processes which are essential in the performance of the desired task. Fourth, different types of items and test formats are desirable for determining how readily students transfer learning of novel tasks and what types of solution strategies they employ. Finally, diverse methods of scoring (beyond number correct and response latency) are needed to diagnose learning difficulties. Thus, cognitive psychology promotes forms of assessment beyond conventional paper-pencil tests relying on unconstructed responses.

Human abilities research has provided the explicit implications for the assessment of school achievement that are highlighted below:

- Identifying component processes provides diagnostic information that can be used to improve performance (Sternberg, 1985).
- By acknowledging the presence of a developmental continuum which underlies human abilities, it is easier to identify the range of children's capabilities and limitations at any given age. The acknowledgement of children's range of capabilities aids in selecting appropriate methods to best assess school achievement (Davidson, 1990).
- Crystallised abilities are highly correlated with student learning from instruction. Prior knowledge is a critical attribute to assess when examining school achievement (Snow, 1982).
- Different response formats should be used, including foils and prompts, to discern the solution strategy employed. Constructed responses may provide critical information that is not available in a simple count of correct responses or of response latency (Kyllonen, 1984).

The purpose of this chapter is to consider the implications of cognitive psychology for the assessment of school achievement. By briefly touching upon the research on human abilities and information processing, it identifies the foundation laid by cognitive psychologists for measuring school achievement.

The impact of cognitive psychology on the measurement of school achievement

It is helpful to distinguish achievements from aptitudes. Snow and Lohman (1989) identify achievement areas, including: language comprehension, both verbal and reading; general knowledge structures; specialised knowledge; and problem solving. These areas are clearly different from special aptitudes such as perception, memory, and attention as well as general aptitudes such as reasoning, fluid-analytic and visual-spatial abilities.

Declarative and procedural knowledge, and personal theories

Some of the most productive research has explored the role of declarative and procedural knowledge in the recall and organisation of information. Cognitive psychologists have also identified the role of schemata and the importance of personal, tacit and naive theories of belief for knowledge organisation and recall.

Declarative knowledge

Declarative knowledge can be defined as a semantic network of facts and ideas in memory (Rumelhart *et al.*, 1986). The relations among the facts and ideas, or knowledge units, are labelled and organised by the learner. Knowledge is usable when the particular nodes in the network are activated. The capacity to recall information is dependent on the way the information is organised. Achievement tests need to estimate the amount and type of organisation a student has developed.

Research studies have focused on several aspects of declarative knowledge organisation. Some studies relate concepts in a particular domain, for example, researchers have studied the way students organise scientific concepts, such as shape, size, number, weight, density, hardness, transparency, solid and liquid (Novak and Musconda, 1991). A second area of research has been the influence of structured prior knowledge and personal belief systems on the acquisition of new knowledge (Carey, 1985; West and Pines, 1985). Other studies have compared differences in levels of expertise among students; most studies of expert learners have been carried out in the sciences and in mathematics. Another area of research has focused on the impact of relevant instruction to change existing knowledge structures, and studies have shown that a student's knowledge structure can be modified through instruction to reflect a textbook's organisation or the instructor's conceptual organisation (Geeslin and Shavelson, 1975; Naveh-Benjamin *et al.*, 1986). The degree of modification of a student's knowledge structure can depend on factors such as the learner's ability to solve problems, as well as the degree of similarity between the student and the instructor's network of declarative knowledge (Thro, 1978; Naveh-Benjamin *et al.*, 1986). Most successful learners personalise knowledge so that it is meaningful to them; it then becomes easier to recall.

These areas of research point to the importance of assessing the organisation of declarative knowledge. It is important to determine how complex the semantic network is and whether the concepts, facts and relations are accurately represented. Prior knowledge plays a critical role; if it is incorrect, it can impede the long-term acquisition of new knowledge. Declarative knowledge structures are different for novice versus expert learners, and these differences can be assessed. Declarative knowledge is sensitive to instruction. It is crucial to the study of achievement, but it must be assessed along with the organisational skills and strategies used to produce it.

Procedural knowledge

Procedural knowledge serves to organise declarative knowledge. It is made up of metacognitive or other complex systems that organise declarative knowledge. Anderson (1983) proposed a three-stage model that describes how skills become proceduralised. In Stage 1, knowledge is declarative and must be consciously worked on using procedural knowledge. Stage 2 is characterised by the use of specific production rules established through large amounts of practice with feedback. As the student practices more, separately learned procedures can be executed sequentially and eventually became a single, smooth action. Finally, in Stage 3 the range of applications may be regarded as very specialised. The procedure is automatised and the learner follows through without conscious processing. This three-stage model illustrates a common view among cognitive psychologists, namely that learning is the process of developing and executing certain production rules (Snow and Lohman, 1989).

Metacognitive processes are essential for proceduralising knowledge. They are used to plan, activate, monitor, evaluate and modify lower order processes (Brown, 1978). Wagner and Sternberg (1984) concluded that they are among the most transferable mental skills.

One of the most important dimensions of procedural knowledge to assess is speed of execution or degree of automaticity. The degree to which cognitive performance deteriorates when there are simultaneous tasks to process is also a means of estimating the degree of proceduralisation of knowledge. Procedural knowledge is critical to expert performance; and students become more expert, their knowledge becomes more proceduralised.

Schemata

"Schemata" is the label applied to the way in which components of declarative and procedural knowledge are organised. Schemata are higher order structures. They can be networks of related concepts, facts and ideas, such as those used in declarative knowledge, or they can be related series of production rules, such as those employed in procedural knowledge, or they can be personal theories or belief systems. Schemata are organised and reorganised through experience and the accumulation of new information. They help individuals understand their experiences; for example, when an individual reads a text, he uses schemata to anticipate the storyline or plot. The schemata that are employed to help in interpretation can vary drastically depending on the reader's expectations. What individuals learn from a text has been shown to be directly related to their expectations and the schemata that were activated in comprehending the content (Anderson and Pearson, 1984). Snow and Lohman (1989) recommend that reading comprehension achievement tests include items that identify appropriate schemata, as well as the student's flexibility in changing them.

Personal theories

Personal theories represent an individual's system of beliefs regarding phenomena. For example, in the area of science learning, students hold personal theories about the concept of gravity. Researchers have been especially intrigued with invalid theories – cases in which an individual's personal theory does not match the publicly held theory. For example, an invalid personal theory about the particulate nature of matter exists when a student believes that molecules of solids do not move and that there is no space between the molecules (Novak and Musconda, 1991). Novick and Nussbaum (1978) helped establish that many learning difficulties develop because students hold misconceptions that are tenaciously maintained even when teachers repeatedly correct them. Ausubel (1968) stated that preconceptions and misconceptions are strongly held and hard to modify. A personally constructed belief system is more difficult to change than a theory or framework presented by the teacher (Driver and Easley, 1978; White and Tisher, 1986). When personal theories change they are more likely to be dropped completely rather than subtly modified (Champagne *et al.*, 1985). Although many researchers recognise the importance of personal theories and their impact on instruction, they are hard to assess. As conventional methods of assessment are inappropriate, a variety of innovative methods have been employed, including: clinical interviews using different amounts of structure, hierarchical schemes for scoring open-ended responses, phenomenography (Marton, 1981), concept-mapping (Novak and Musconda, 1991), word association (Geeslin and Shavelson, 1975), and analysis of verbal protocols (Greeno, 1978). To date, researchers have had some success in detecting personal theories and knowledge structures in research studies. However, these methods have not been successfully applied in any large-scale testing programmes.

The critical role of prior knowledge

Knowledge is acquired over time, and prior knowledge grows in size and complexity with the normal development of fluid and crystallised intelligence. Personal theories about concepts and the interrelationships among them may be reflected in student's prior knowledge. Thus, prior knowledge can contain correct or incorrect content. Moreover, prior knowledge units can be linked to one another in a meaningful or, conversely, in an arbitrary manner. "Thus, prior knowledge is often partial, incomplete, or incorrect in idiosyncratic ways" (Snow and Lohman, 1989, p. 304). Newly acquired information may make aspects of prior knowledge irrelevant to the learner, and prior knowledge may, at times, have to be modified or replaced by new knowledge. Moreover, newly acquired knowledge may serve to change faculty linkages or to build new connections among knowledge units. Regardless of the accuracy or meaningfulness of its content, the amount of prior knowledge a student possesses and the way the knowledge is organised may serve to facilitate or impede achievement. Recall of information depends on how well it is organised and how easily it can be accessed.

Measuring school achievement in content areas

Problem-solving strategies specific to particular characteristics of the knowledge domain have been identified. Cognitive psychologists have identified paths of expertise in particular knowledge domains.

Novice and expert problem solvers

Expertise is specific to the particular characteristics of knowledge domains. Important distinctions have been made between expert and novice problem solvers by focusing on the ways they represent problems and arrive at solutions (Hayes, 1981). Novices and experts represent problems differently. They differ in the complexity and extent of their knowledge schemata, and in the types of search strategies they use for solving problems (Chi *et al.*, 1981). Table 11.1 presents some of the differences between novice and expert problem solvers.

Expertise is specific to content areas and the mechanisms by which problems are solved in different content areas have been documented. For example, using interviews, Larkin and Reif (1979) established differences in strategies used by novice and expert problem solvers in physics. Novices quickly resorted to converting the physics problem to a series of equations and immediately tried to manipulate the equations. Experts examined the terms of the problem more thoroughly and clarified that their position was sound before introducing the equations. They dealt with the actual mathematics at the very end of the problem-solving situation. Experts spend more time working on the critical elements of a problem before seeking at a solution. Experts and novices also differ in basic metacognitive processes. Experts are "vigilant managers" who monitor problem-solving performance and strive for efficiency and accuracy. Novices are not "vigilant managers" and may allocate all their time to an aspect of a problem that does not bring them closer to the solution (Schoenfeld, 1983). Expert-novice differences have also been established in biology, chemistry, mathematics, history and other social sciences (Haertel and Gerlach-Downie, 1992). Table 11.1 draws implications for assessment based on differences between novice versus expert performance.

Problem-solving strategies in curricular areas

Curricular domains differ in the way problems are structured. Some domains, such as mathematics, have very well-defined and well-structured problems. On the other hand, instruction in the social sciences is often conducted using ill-structured problems with little specification. Well-structured problems have a clear initial state, a well understood set of permissible operations and a clear end state.

Most cognitive research has been done in curricular areas with well-specified problems such as the sciences and mathematics. Assessment differs in areas with well versus ill-structured problems. It is easier to diagnose faulty algorithms and misconceptions when the curricular content and problem-solving procedures are explicit. If many errors are rule-governed, as many cognitive psychologists believe, then errors are more likely to be diagnosed in well-structured curricular areas. A student makes an error because he or she employs an incorrect algorithm or rule – the error is not the result of failing to learn, but the result of learning the wrong rule (Resnick, 1982). Brown and Burton (1978) and Brown and Van Lehn (1982) have illustrated a particular approach to error analysis – the identification of "buggy" algorithms. Brown and his colleagues did not simply wish to identify common buggy algorithms, rather, they developed a generative theory of bugs that account for how they are acquired. They have developed a minimum set of diagnostic exercises or tests that can distinguish all identified bugs, and the results of the tests are used to remediate student performance (Romberg and Carpenter, 1986).

Table 11.1 **Differences in the ways expert and novice performers organise and monitor problem-solving tasks**

Nature of task	Characteristics of expert performance	Characteristics of novice performance	Assessment implications
Organising the problem	Tend to organise knowledge based on higher-order principles. Try to put structure into information as they learn it. Have a complex and very complete knowledge schemata.	Store knowledge in more isolated bits or sort it on the basis of surface characteristic. Do not try actively to structure information as they are exposed to it. Knowledge schemata is less well organised.	Assess the complexity of the knowledge structure. Determine the conceptual foundations and principles that underlie the skills.
Problem-solving	Solve problems using fundamental principles. Attend to a problem's structure.	When solving a problem attend to superficial causes and the surface details of the problem.	Distinguish chaotic and flexible reasoning.
Monitoring the problem solution	Carefully monitor problem-solving performance. Generate rich problem representations as a guide to solution and search.	Do not monitor problem-solving performance. Knowledge and skills can be wasted. Do not modify the problem-solving approaches even if new information is acquired.	Determine the type of search strategy used: random or exhaustive search, a means-end analysis, or generate and test a variety of solutions.

There have also been studies of how students solve story problems. There are differences in individual's capability to arrive at and implement solution strategies. Mayer (1985) identified five kinds of information needed for problem solution: linguistic knowledge (parsing sentences, word meanings); factual knowledge (crystallised knowledge such as the decimal system); schema knowledge for problem types (combinations or distance-rate-time problems); strategic knowledge of how to develop, monitor and revise a solution plan; and algorithmic knowledge (division or subtraction). Mayer has also identified two phases in problem-solving planning and execution.

Diagnosis of errors in problem solving

Cognitive psychologists have documented that a variety of types of errors are made during problem solving (Resnick, 1976; Snow and Madinach, 1991). Some occur across all types of problem solving situations, for example, errors in the planning and execution of solution strategies. These general errors are often due to the poor organisation of an individual's semantic networks. Others, in contrast, occur within specific content domains; for example, buggy algorithms in mathematics and invalid theories of natural phenomena in science. Some of these content-specific errors may occur consistently and reflect a faulty model, theory or algorithm. Or they may occur randomly or without any apparent pattern. Other errors occur because particular skills are inappropriately applied. Morever, some errors are common and are evidenced in the thinking of many individuals, while others are rare and occur in only a few students. The latter may be attributed to students' personal theories, idiosyncratic networks of knowledge units or mislearned content. Assessment of achievement should diagnose the type and source of errors that students make.

Alternative routes to achievement

Cognitive psychologists have documented the fact that a student might take a multitude of routes to solve a problem. Some are those presented during classroom instruction or in a textbook illustration. Others are highly idiosyncratic and reflect the unique and complex knowledge structures individuals develop over their lifetime. Assessments need to be sensitive to alternative routes, and the value of alternative solution strategies needs to be assessed in terms of the students' developing long-term understanding and knowledge structures that facilitate learning.

Implications of cognitive psychology for improved achievement tests

- Declarative knowledge is a semantic network of knowledge units that encompasses facts, ideas and relations among concepts. The organisation of declarative knowledge is directly related to a student's ability to retrieve and recall information. Assessment should measure the amount of declarative knowledge a student possesses and identify the type of organisation used to store it (Rumelhart *et al.*, 1986; Snow and Lohman, 1989).
- Procedural knowledge serves to structure declarative knowledge. It encompasses higher order skills and strategies including metacognitive skills. As knowledge becomes proceduralised, its processing increases in automaticity. Proceduralised knowledge does not demand conscious attention on the part of the student. Assessments should be designed to measure the speed of execution of cognitive

tasks. Attentional demands may be manipulated by introducing concurrent tasks or increasing the processing load of the assessment procedure (Anderson, 1983).

- Prior knowledge encompasses all the facts and relations, as well as other knowledge units, available to the learner. As a child develops, his or her knowledge units increase in number, and particular knowledge units can be modified, connected, transformed or completely replaced. Memory serves to help organise prior knowledge. Assessment should be designed to be sensitive to the learner's developmental sequences and stages of understanding.
- The role of feedback is critical for proceduralising declarative knowledge. Assessment, by its very nature, provides feedback. Supplying students with assessment results that are immediate and that focus attention on inaccuracies in process or content is essential for students to move towards an advanced stage of proceduralisation (Anderson, 1983).
- Schemata are higher order structures. Some are publicly affirmed and others can be seen as personal belief systems. Personal theories may vary in comprehensiveness, explaining isolated concepts or covering a wide range of events. Personal theories are especially resistant to change. When they do change, however, they typically are altered dramatically, or dropped completely, rather than modified in subtle ways. Assessments should determine the appropriateness, comprehensiveness, and flexibility of the schemata through error analysis (Glaser *et al.*, 1984).
- There are a variety of errors learners make during problem-solving. Types of errors include "buggy" algorithms in mathematics, application of naive theories and misconceptions in the sciences. Other errors reflect overgeneralisations or undergeneralisations of particular skills. Some errors may reflect unique factors exemplified by personal theories, idiosyncratic networks of knowledge units or mislearned content. Assessments should diagnose the type and source of errors. Errors should be analysed in terms of: degree of incorrectness; specificity to content domain; and consistency of occurrence. Idiosyncratic responses provide critical information for use in diagnosing instructional deficiencies (Brown and Burton, 1978; Champagne *et al.*, 1980).
- Problem-solving strategies are specific to particular content areas. Therefore, assessment of problem-solving must be domain-specific. Different curricular areas use different types of problems for instruction. Problems can be well-structured, ill-structured or wholly unspecified. Assessment of these different types of problems requires a broad range of techniques, including: novel response formats, use of constructed responses and alternative scoring techniques (Snow and Lohman, 1989).
- All students use solution strategies in solving problems. Both the successful planning and executing of the solution strategies is required. Students use alternative solution strategies on the basis of their own knowledge structures and the nature of the particular curricular areas. Assessment should identify the alternative solution strategies, as well as the capacity of the student to change strategies if unsuccessful (Hayes, 1981).
- Metacognitive skills co-ordinate lower order cognitive processes. They are responsible for planning, activating, monitoring, evaluating and modifying lower order skills. Metacognitive skills are believed to be transferable across content areas. Assessments need to determine the kinds of metacognitive skills that a student uses – those that are helpful and those that fail to help. General as well as content-specific metacognitive skills should be identified.

- Differences between expert and novice performance provide detailed information about the nature of expertise in a specific domain. Assessment should be based on how experts represent the problem, how they plan and select a solution strategy and how they execute that strategy (Chi *et al.*, 1981).
- Diagnostic testing is a key to improving student performance. By identifying the source and type of errors a student makes, teachers can provide feedback and improve the individual student's performance. Diagnostic testing can provide information on such factors as: the quality of organisation of semantic networks, the degree of automaticity in cognitive skills, the presence of errors or misconceptions in personal belief systems or the choice of poor solution strategies in problem-solving. There are many more types of error analysis that could provide important data on a student's achievement. Diagnostic testing is an essential component of adaptive instruction, which links a student's test performance to instruction, curricular activities and remediation (Wang and Lindvall, 1984; Corno and Snow, 1986).
- The components of an adaptive test battery to measure students' expertise in a given content area have been described by Snow and Lohman (1989) and include the following three elements: assess abilities such as reading comprehension and verbal abilities; assess the nature and type of the general knowledge schemata and determine whether they impede or facilitate learning; and assess the declarative knowledge structures in the particular content area (*e.g.* multiplication tables or vocabulary). Upon completion of this battery a teacher could determine where the student fits in a particular instructional unit.

New directions in research on assessing achievement

Current research efforts are examining the role cognitive psychology can play in developing new types of assessments. For example, the novice versus expert distinction has spawned active research in many content areas, and more information is being acquired that describes differences in problem-solving between these two groups. These differences provide fruitful implications for assessment of achievement, as does the use of faceted achievement tests, where comparisons can be made, for example, among different questioning modes or levels of topic difficulty. Faceted achievement tests make it possible to determine the impact of the test or item format on student performance. For example, researchers could determine how much a specific type of test format affects students of different abilities, ages, and instructional histories (Calfee and Drum, 1979).

Another important line of work is the use of dynamic assessment and applications of Vygotsky's zone of proximal development (Vygotsky, 1978). As stated earlier in this chapter, expertise is always specific to a knowledge domain. This is reflected in the shift on the part of educators towards curriculum-based assessment and away from more general indices of ability, aptitude or achievement (Keating, 1990). Educators need specific guidance on how to proceed – hence, the need for adaptive instruction. Assessments must be grounded in strong cognitive theory and yet closely linked to instruction. Because students move along a path of increasing expertise, a dynamic form of assessment is required. What is needed is not just an assessment of what a student can do independently at a given moment in time, but how far he or she can proceed with increasing levels of assistance or scaffolding. This range of performance is referred to by

Vygotsky as the range of proximal development. Dynamic assessment is based on the idea of increasing levels of expert performance and could fit into an adaptive instruction framework (see also Ivic, Chapter 10, this volume).

Conclusion

There is an increasing body of research on cognition that can be used to inform assessment design. There are reasons, however, to be cautious in applying these new findings. First, there is little information about the statistical characteristics of these new forms of assessment. New standards and rules of evidence may be required to examine the reliability and validity of scores from tests that have very different formats from those of the more widely used multiple-choice achievement tests. These new forms of assessment, if they are to be used for monitoring and diagnosing individual student progress or for group accountability, will need to demonstrate some form of representativeness, comparability, consistency and generality (Linn, 1991). Technical standards can be established both by the use of expert judgement and statistical analyses. However, recent results of the 14-nation student writing assessment attest to the difficulties of using expert ratings. Alan Purves, director of the international writing assessment study, stated that the study had been an "interesting failure, which should make us more modest about writing assessment". Purves also stated that the use of expert ratings to judge the quality of the students' drafts had been troublesome and that "those advocating the use of performance assessments in writing and other subjects 'need to be more honest' about problems in comparing ratings" (Rothman, 1991).

In addition to the variety of technical problems that have to be resolved before the newer forms of assessment can be used without apprehension, their cost also presents a challenge. Developing the types of tests described in the earlier section of this chapter, with novel formats, constructed responses and diagnostic capabilities, requires more resources than the development of standardised, machine-scorable, multiple-choice achievement tests. Shavelson (1991) has described "rhetoric versus reality" in a project devoted to developing "authentic" assessments for hands-on mathematics and science instruction. He presents good and bad news about "authentic" assessment and, in particular, he reports some striking findings showing that test format or method of assessment can cause large differences in student scores.

Some advocates of new forms of assessment assert that the synergy generated during the development and implementation of new assessment forms can produce desirable benefits. The California Assessment Program has reported that teachers involved in the development and scoring of new assessments find the process to be the most effective staff development effort in which they had ever participated (Carlson, 1991). The cost of developing, administering and scoring the newer forms of assessment must be determined.

If new forms of assessment are to be widely embraced, it is necessary to ensure that the promise of diagnosis can be fulfilled. To make decisions about the education of individual children on the basis of assessments whose validity has yet to be well understood is troublesome. In addition, group accountability measures are critical benchmarks used by nations, school districts and local schools to determine whether they are achieving the goals they set. The assessments used for these comparisons must be valid, reliable, and cost-efficient and they must not place a burden on the test-taker. These concerns must be weighed against the promise of the new assessments based on a rich foundation of cognitive psychology findings.

Bibliography

ANDERSON, J.R. (1983), *The Architecture of Cognition*, Harvard University Press, Cambridge, MA.

ANDERSON, R.C. and PEARSON, P.D. (1984), ''A Schema-theoretic View of Basic Processes in Reading'', in Pearson, P.D., Bar, R., Kamil, M.L. and Mosenthal, P. (eds.), *Handbook of Reading Research*, Longman, New York, pp. 255-292.

AUSUBEL, D.P. (1968), *Educational Psychology*, Holt, Rinehart and Winston, New York.

BINET, A. and SIMON, T. (1905), ''Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux'', *L'année psychologique*, Vol. 11, pp. 236-245.

BORING, E.G. (1929), *A History of Experimental Psychology* (2nd edition), Appleton-Century-Crofts, New York.

BROWN, A.L. (1978), ''Knowing When, Where, and How to Remember'', in Glaser, R. (ed.), *Advances in Instructional Psychology*, Vol. 1, Lawrence Erlbaum, Hillsdale, NJ, pp. 77-165.

BROWN, J.S. and BURTON, R.R. (1978), ''Diagnostic Models for Procedural Bugs in Basic Mathematical Skills'', *Cognitive Science*, Vol. 2, pp. 155-192.

BROWN, J.S. and van LEHN, K. (1982), ''Towards a Generative Theory of 'Bugs''', in Carpenter, T.P., Moser, J.M., and Romberg, T.A. (eds.), *Addition and Subtraction: A Cognitive Perspective*, Lawrence Erlbaum, Hillsdale, NJ, pp. 117-135.

BURT, C. (1940), *The Factors of the Mind*, University of London Press, London.

CALFEE, R.C. and DRUM, P.A. (1979), ''How the Researcher Can Help the Reading Teacher with Classroom Assessment'', in Resnick, L.B. and Weaver, P.A. (eds.), *Theory and Practice of Early Reading*, Vol. 2, Lawrence Erlbaum, Hillsdale, NJ, pp. 173-205.

CAREY, S. (1985), *Conceptual Change in Childhood*, MIT Press, Cambridge, MA.

CARLSON, D. (1991), ''Changing the Face of Testing in California'', in *California Curriculum News Report*, Vol. 16 (3), California Association of County Superintendents of Schools, San Francisco, CA.

CARROLL, J.B (1985), ''Domains of Cognitive Ability'', paper presented at the meeting of the American Association for the Advancement of Science, Los Angeles, May.

CHAMPAGNE, A.B., GUNSTONE, R.F. and KLOPFER, L.E. (1985), ''Instructional Consequences of Students' Knowledge about Physical Phenomena'', in West, L.H.T. and Pines, A.L. (eds.), *Cognitive Structure and Conceptual Change*, Academic Press, New York, pp. 61-90.

CHAMPAGNE, A.B., KLOPFER, L.E. and ANDERSON, J.H. (1980), ''Factors Influencing the Learning of Classical Mechanics'', *American Journal of Physics*, Vol. 48, pp. 1074-1079.

CHI, M.T.H., FELTOVICH, P. and GLASER, R. (1981), ''Categorization and Representation of Physics Problems by Experts and Novices'', *Cognitive Science*, Vol. 5, pp. 121-152.

CORNO, L. and SNOW, R.E. (1986), "Adapting Teaching to Individual Differences Among Learners", in Wittrock, M.C. (ed.), *Handbook of Research on Teaching* (3rd edition), Macmillan, New York, pp. 605-629.

CSIKSZENTMIHALYI, M. and ROBINSON, R. (1986), "Culture, Time and the Development of Talent", in Sternberg, R. (ed.), *Conceptions of Giftedness*, Cambridge University Press, Cambridge, pp. 264-284.

DAVIDSON, E. (1990), "Intelligence Recreated", *Educational Psychologist*, Vol. 25 (3-4), pp. 337-354.

DILLON, R.F. (ed.) (1985), *Individual Differences in Cognition*, Vol. 2, Academic Press, New York.

DRIVER, R. and EASLEY, J. (1978), "Pupils and Paradigms: A Review of Literature Related to Concept Development in Adolescent Science Students", *Studies in Science Education*, Vol. 5, pp. 61-64.

FINN, JR., C.E. (1991), *We Must Take Charge: Our Schools and Our Future*, The Free Press, New York.

GARDNER, H. (1983), *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books, New York.

GEESLIN, W.E. and SHAVELSON, R.J. (1975), "An Exploratory Analysis of the Representation of a Mathematical Structure in Students' Cognitive Structures", *American Educational Research Journal*, Vol. 12, pp. 21-39.

GLASER, R. (1984), "Education and Thinking: The Role of Knowledge", *American Psychologist*, Vol. 39, pp. 93-104.

GLASER, R., LESGOLD, A. and LAJOIE, S. (1984), "Towards a Cognitive Theory for the Measurement of Achievement", in Ronning, R., Glover, J., Conoley, J.C. and Witt, J. (eds.), *The Influence of Cognitive Psychology on Testing and Measurement: The Buros-Nebraska Symposium on Measurement and Testing*, Vol. 3, Lawrence Erlbaum, Hillsdale, NJ.

GREENO, J.G. (1978), "A Study of Problem Solving", in Glaser, R. (ed.), *Advances in Instructional Psychology*, Vol. 3, Lawrence Erlbaum, Hillsdale, NJ, pp. 13-75.

GUILFORD, J.P. (1967), *The Nature of Human Intelligence*, McGraw-Hill, New York.

GUSTAFSSON, J.E. (1984), "A Unifying Model for the Structure of Intellectual Abilities", *Intelligence*, Vol. 8, pp. 179-203.

HAERTEL, G.D. and GERLACH-DOWNIE, S. (1992), "Cognitive Psychology: A Foundation for Improved Student Assessment", paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April.

HAYES, J.R. (1981), *The Complete Problem Solver*, Franklin Institute Press, Philadelphia.

KEATING, D. (1990), "Charting Pathways to the Development of Expertise", *Educational Psychologist*, Vol. 25 (3-4), pp. 243-268.

KORNHABER, M., KRECHEVSKY, M. and GARDNER, H. (1990), "Engaging Intelligence", *Educational Psychologist*, Vol. 25 (3 - 4), pp. 177-200.

KYLLONEN, P.C. (1984), "Information Processing Analysis of Spatial Ability", Doctoral dissertation, Stanford University, *Dissertation Abstracts International*, 45, 819A.

LARKIN, J.H. and REIF, F. (1979), "Understanding and Teaching Problem Solving in Physics", *European Journal of Science Education*, Vol. 1, pp. 191-203.

LINN, R.L. (1991), "Dimensions of Thinking: Implications for Testing", in Idol, L. and Jones, B.F. (eds.), *Education Values of Cognitive Instruction: Implications for Reform*, Lawrence Erlbaum, Hillsdale, NJ, pp. 179-208.

MARTON, F. (1981), "Phenomenography: Describing Conceptions of the World Around Us", *Instructional Science*, Vol. 10, pp. 177-200.

MAYER, R.E. (1985), "Mathematical Ability", in Sternberg, R.J. (ed.), *Human Abilities: An Information Processing Approach*, Freeman, San Francisco, pp. 127-150.

MICHEL, C. (1990), "Information-processing Theory", in Thomas, R.M. (ed.), *The Encyclopedia of Human Development and Education*, Pergamon Press, Oxford, pp. 78-82.

NAVEH-BENJAMIN, M., McKEACHIE, W.J., LIN, Y-G. and TUCHER, D.G. (1986), "Inferring Students' Cognitive Structures and their Development Using the Ordered Tree Technique", *Journal of Educational Psychology*, Vol. 78, pp. 130-140.

NOVAK, J.D. and MUSCONDA, D. (1991), "A Twelve Year Longitudinal Study of Science Concept Learning", *American Educational Research Journal*, Vol. 28 (1), pp. 117-154.

NOVICK, S. and NUSSBAUM, J. (1978), "Junior High School Pupils' Understanding of the Particulate Nature of Matter: An Interview Study", *Science Education*, Vol. 62, pp. 273-281.

PIAGET, J. (1963), *The Psychology of Intelligence*, International Universities Press, New York.

RESNICK, L.B. (1976), "Task Analysis in Instructional Design: Some Cases from Mathematics", in Klahr, D. (ed.), *Cognition and Instruction*, Lawrence Erlbaum, Hillsdale, NJ, pp. 51-80.

RESNICK, L.B. (1982), "Syntax and Semantics in Learning to Subtract", in Carpenter, T.P., Moser, J.M. and Romberg, T.A. (eds.), *Addition and Subtraction: A Cognitive Perspective*, Lawrence Erlbaum, Hillsdale, NJ, pp. 136-155.

ROMBERG, T.A. and CARPENTER, T.P. (1986), "Research on Teaching and Learning Mathematics: Two Disciplines of Scientific Inquiry", in Wittrock, M.C. (ed.), *Handbook of Research on Teaching* (3rd edition), Macmillan, New York, pp. 850-873.

ROTHMAN, R. (1991), "14-nation Student Writing Assessment Called Failure", *Education Week*, 17 April.

RUMELHART, D.E., McCLELLAND, J.L. and the PDP Research Group (1986), *Parallel Distributed Processing Foundations*, MIT Press, Cambridge, MA.

SCHOENFELD, A.H. (1983), "The Wild, Wild, Wild, Wild, Wild World of Problem Solving", *For the Learning of Mathematics*, Vol. 3, pp. 40-47.

SHAVELSON, R.J. (1991), "Authentic Assessment: The Rhetoric and the Reality", paper presented at the Annual Meeting of the American Educational Research Association, Chicago, May.

SNOW, R.E. (1982), "Education and Intelligence", in Sternberg, R.J. (ed.), *Handbook of Human Intelligence*, Cambridge University Press., New York, pp. 493-585.

SNOW, R.E. (1986), "Individual Differences and the Design of Educational Programs", *American Psychologist*, Vol. 41 (10), pp. 1029-1039.

SNOW, R.E. and LOHMAN, D.F. (1989), "Implications of Cognitive Psychology for Educational Measurement", in Linn, R.L. (ed.), *Educational Measurement*, Vol. 3, Ace and Macmillan, New York.

SNOW, R.E. and MADINACH, E.B. (1991), *Integrating Assessment and Instruction: A Research and Development Agenda* (Research Report RR-91-8, February), Educational Testing Service, Princeton, NJ.

SPEARMAN, C.E. (1927), *The Abilities of Man*, Macmillan, London.

STERNBERG, R.J. (1985), *Beyond IQ: A Triarchic Framework for Intelligence*, Cambridge University Press, New York.

STERNBERG, R.J. (1990), "T and T is an Explosive Combination: Technology and Testing", *Educational Psychologist*, Vol. 25 (3-4), pp. 201-222.

THOMAS, R.M. (1985), *Comparing Theories of Child Development*, Wadworth, Belmont, CA.

THRO, M.P. (1978), "Relationships between Associative Patterns and Content of Physics Concepts", *Journal of Educational Psychology*, Vol. 70, pp. 971-978.

THURSTONE, L.L. (1938), "The Perceptual Factor", *Psychometrika*, Vol. 3, pp. 1-12.

VERNON, P.E. (1971), *The Structure of Human Abilities*, Methuen, London.

VERNON, P.E. (1990), "The Use of Biological Measures to Estimate Behavioral Intelligence", *Educational Psychologist*, Vol. 25 (3-4), pp. 293-305.

VYGOTSKY, L.S. (1978), *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press, Cambridge, MA.

WAGNER, R.K. and STERNBERG, R.J. (1984), "Alternative Conceptions of Intelligence and their Implications for Education", *Review of Educational Research*, Vol. 25, pp. 179-224.

WANG, M.C. and LINDVALL, C.M. (1984), "Individual Differences and School Learning Environments", *Review of Research in Education*, Vol. 11, pp. 161-225.

WEST, L. and PINES, L. (eds.) (1985), *Cognitive Structure and Conceptual Change*, Academic Press, New York.

WHITE, R.T. and TISHER, R.P. (1986), "Research on Natural Sciences", in Wittrock, M.C. (ed.), *Handbook of Research on Teaching* (3rd edition), Macmillan, New York.

Chapter 12

Achievements of Learning

by

Patricia Broadfoot
University of Bristol, United Kingdom

Changing work practices, technological advances and a demographic decline in the absolute numbers of young people in many advanced industrialised societies highlight the increasingly urgent need for students to be equipped with the skills and qualities that will enable them to adapt to a variety of occupational roles. Problem-solving ability, personal effectiveness, thinking skills and willingness to accept change are typical of the general competencies that are being sought in young people. Although aspects of such competencies are embodied in traditional subject syllabuses, the tension between pressures for curriculum development to reflect the acknowledged need to provide students with learning opportunities in which they can develop these general competencies, on the one hand, and the inability of the assessment industry, as yet, to give widespread support to such developments, on the other, is becoming increasingly marked (Frederiksen and Collins, 1989). Yet, as Nickerson (1989) points out, until such assessment procedures are available, not only will it be impossible to produce valid judgements about the success of the educational enterprise in terms of achieving its goals, it will also give rise to a situation where certain desired educational outcomes will be neglected.

Thus, the challenge facing those charged with the evaluation of education systems is to find ways of developing and measuring indicators that adequately reflect the full range of educational goals. Not to do so will result in inadequate information about performance and a tendency to ignore certain teaching objectives.

What follows is an attempt to consider some of the problems associated with the assessment of educational outcomes; to review work on the mapping of learning goals in ways that cut across specific programmes of instruction; to explore some of the more novel assessment instruments that are currently being pioneered; and, finally, to consider these issues in relation to strategies for conducting national and international surveys of student learning outcomes.

Limitations of traditional assessment

Criticisms of traditional performance testing

Historically, the practice of educational assessment has been largely driven by a perceived need to measure individual capacity. Associated with this has been the desire to measure individual differences as the basis for fair and appropriate selection for different opportunities and roles within society. Thus the psychometric tradition of assessment has predominated and with it has come an overwhelming emphasis on the need for reliability in test application, so that assessment may be seen to operate fairly and consistently in determining chances in life. The question of validity – whether the test does indeed measure what it is intended to measure – has arguably been subordinated to the overwhelming need for comparability of results. The preoccupation with reliability has tended to lead to a concentration on what is more easily measurable, such as knowledge and understanding, and a relative neglect of higher-level intellectual skills such as "thinking" and of those affective qualities which are crucial to learning, but which are much more difficult to measure using psychometric techniques. Whilst higher-order skills and personal qualities have been the subject of informal research studies that testify to their perceived importance, much of this research has tended to be intuitive, impressionistic and explicitly subjective. More systematic attempts to incorporate the assessment of high-level learning outcomes in, for example, reviews of education systems' performance have typically foundered because of problems of conceptualisation and the lack of acceptable assessment techniques (Assessment of Performance Unit, 1984; Shepard, 1990).

However, dissatisfaction with traditional performance testing in education is spreading. Criticisms concern the emphasis on comparison between students rather than describing specific and changing levels of attainment; the frequent mismatch between curriculum and test content; the pressure to test in just a few aspects of a programme of instruction; the assumption that students must learn and be assessed on the "basics" before going on to more complex intellectual processes; and the methodology commonly employed in student assessment. Another fundamental reason for dissatisfaction is that the tests can be shown not to do what they aspire to do: namely, to provide objective, reliable evidence on the cognitive aspects being measured (Ingenkamp, 1977; Lee-Smith, 1990; Raven, 1989).

The "backwash" effect of testing

A further criticism of traditional testing approaches concerns their undesirable backwash effect on instruction. It is also increasingly being recognised that the gap between what is covered by formal assessment and the desired learning outcomes that are not tested has major repercussions on the emphasis given in what is taught.

Thus, some theoreticians are now arguing for a reinterpretation of construct validity that will take explicit account of this backwash effect. Starting from the premise that test validity must encompass the actual and potential consequences of test score interpretation and use, it can be argued that a test should be regarded as "systematically" valid when it induces in the education system curricular and instructional changes that foster the development of the (cognitive) skills that the test is designed to measure (Frederiksen and Collins, 1989).

Studies in the field of metacognition reveal just how important assessment is in defining the attitude students take towards their work, their sense of ownership and control, the strategies they employ in learning and their confidence and self-esteem – all factors which impact on the learning achieved. Thus, the search for new assessment approaches has been fuelled by the growing recognition of the limitations of traditional approaches to assessment, especially the need for a greater concern with validity; by concern about the influential but potentially quixotic role of "personal" assessment; by a desire to harness the powerful impact of assessment to promote rather than inhibit learning. Above all, increasingly explicit demands by modern economies both for the encouragement of new kinds of learning outcome and for information about a much broader range of skills and qualities have combined to create a climate in which both the potential outcomes of learning and their realisation or non-realisation are the subject of new definitions and new approaches to assessment.

The link between theory and practice

The fashion for criterion-referenced testing, which is characterised by an emphasis on the assessment of "discrete" competencies that have immediate behavioural outcomes that can be segmented and individually tested, and which can be linked to a specific school curriculum (Cole, 1990, p. 3), has arguably not helped the evolution of assessment techniques to address the more "associative" and "interpretive" uses of learning embodied in more open-ended assignments. Part of the reason for the preoccupation with defining and assessing discrete, linear learning trajectories may be the persistence of outdated theories of learning among psychometricians (Shepard, 1990) and a corresponding failure to assume the implications for assessments of advances in cognitive psychology.

Accordingly, many traditional assessment techniques may be shown to have serious shortcomings. First, their underlying theory of learning often does not reflect important new developments in cognitive psychology. Second, their content does not help to encourage the learning of all the competencies identified as goals for the education system. Third, the assessment procedures used may adversely affect the students' approach to learning and to the classroom climate as a whole. Fourth, the effects of such testing encourage users to think of educational outcomes in terms that are not constructive either as a basis for judging the true achievements of individuals or as a basis for accountability in judging the achievement of the education system as a whole.

It may be concluded that there is a fundamental flaw in much contemporary assessment, which helps to explain the inability of most existing modes of assessment to improve student performance. As Baker *et al.* (1990) argue, assessment approaches are needed that measure significant learning in a way that supports desired performance and that provides reliable information about outcomes. In particular, assessment must address more generic cognitive learning tasks such as "deep" understanding and problem-solving. How these more general goals may be identified and assessed is discussed below.

Attempts to "map" learning outcomes

Curriculum theory-based models

Some early and highly influential attempts to "map" learning domains were undertaken in the realm of curriculum theory. Hirst's (1974) attempt to distinguish, for example, the unique and exclusive forms of knowledge that constitute a "liberal education" identified mathematics, physical sciences, human sciences, history, religion, literature, the fine arts, philosophy, together with three different "fields" of knowledge – the practical, the theoretical, and the moral.

Another project (Her Majesty's Inspectorate, 1977) identified the following dimensions of learning: the physical, the aesthetic and creative, the ethical, the linguistic, the mathematical, the scientific, the social and political, and the spiritual (see also Oates, 1990). Whilst such curricular maps provide a rationale for judging the adequacy of educational outcomes in terms of coverage (on the assumption that all students have some familiarity with each of the areas), they have little to say about the level or form that achievement should take. Thus they are more properly a rationale for judging educational inputs than outcomes.

An alternative approach to mapping the range of learning outcomes is associated with attempts to devise taxonomies of educational outcomes. Such taxonomies have described both the expression of skills within domains in a hierarchical form and the distinction between domains of different kinds of behaviour. The taxonomy of Bloom *et al.* (1956) has been particularly influential. It distinguishes between cognitive, affective and psychomotor domains. But although the hierarchy of cognitive skills Bloom and associates identified (*i.e.* knowledge, comprehension, application, analysis, synthesis and evaluation) has been useful in certain respects, it has also proved difficult to defend the purity of the hierarchical boundaries on which the taxonomy is based. This is especially true of the distinction between lower- and higher- order skills referred to above. Thirty years later, it is becoming apparent that such taxonomies artificially separate the ties that inevitably exist between the domains identified by Bloom and his colleagues. The implication is that the validity of the taxonomy may be in doubt.

Other problems also arise when the cognitive hierarchical model is used. One is that a question may make different demands on students depending on their specific past experiences with a given topic (Crooks, 1989). Although the levels are frequently compressed into three broad categories – recall, application and problem-solving, in which the student is required to transfer existing knowledge and skills to novel situations – a review of the research reveals that the recall level predominates in classroom testing and that "classroom examinations often fail to reflect teachers' stated instructional objectives" (Haertel, 1986, p. 2). Where indirect measures are used to judge higher-order skills, it is in these operations and not in the truly desired learning that children tend to be coached (Izard, 1990).

More recently, Gagne *et al.* (1988) have moved away from models of desired learning outcomes based on curriculum theory and philosophy towards a reformulation of learning goals. This approach is marked by a departure from academic subjects as the context for assessment; it acknowledges that some aspects of learning are promoted across a wide range of different subjects.

Snow (1989) reviews a range of literature on learning theory in order to highlight the central role of the concepts of progression and transferability in learning and points out the different kinds of skills involved in applying existing learning to new situations. He emphasizes too that successful learning depends upon the flexible use by learners of different learning strategies. Also vital is the role of confidence in linking cognitive knowledge, strategy and skill. Motivation and self-regulation emerge as central to successful learning through their emphasis on "action in control", "mindful adaptation" and various other metacognitive skills. Indeed, as Cronbach (1984) argues, it is not without significance that these constructs come close to Binet's original definition of intelligence.

Pragmatic definitions of competence

Other attempts at defining high-level cognitive skills and general learning outcomes which straddle the cognitive, affective and conative divides are much more pragmatic in character and reflect the substantive competencies sought in school-leavers particularly by employers. These forms of learning are often described as "core competencies" or "general skills", and a number of attempts have been made to identify them. Some of these efforts are based on an analysis of the skills and competencies students will need. Others are derived from a range of curriculum subjects.

A clear summary of the different types of competence is provided by the National Curriculum Council of England and Wales (1990), which refers to:

- Fundamental core skills underpinning almost all employment functions and formal programmes of instruction, identified as:
  - problem-solving;
  - communication;
  - personal skills.
- Equally important but less generic core skills which may be selectively present in learning and occupational performance, namely:
  - numeracy;
  - information technology;
  - modern language competence.
- Cross-curricular themes which ought to constitute a part of the learning goals for senior secondary and general vocational education, such as:
  - social and economic understanding;
  - scientific and technological understanding;
  - aesthetic and creative understanding.

A rather similar list is generated as a basis for students' descriptive "records of achievement" by the Department of Education and Science (1989):

- Communication skills:
  These would include writing and speaking, and possibly other means of communication.
- Working with others:
  This could cover a variety of different forms, from formally organised group work in subject lessons, through semi-formal work on group projects, to working with old people or young children.
- Organising work:

This could cover matters such as preparing and planning tasks, organising time and meeting deadlines.

- Information handling:
  This could cover matters such as the ability to find, analyse and present information and to frame and test hypotheses.
- Personal qualities:
  This would include qualities such as reliability and enthusiasm.

These are essentially *de facto*, common sense descriptions of the different competencies that are or should be demonstrated by successful learners. As such they cut across domain distinctions and mix in ways that makes their assessment deeply problematic. Attitudes, skills and knowledge (for example in the "core competency" of information technology); transient as well as relatively permanent personal characteristics (motivation as opposed to ability to work in a group, for example); content that is likely to be explicitly taught, such as health education; and content that is likely to be largely among the implicit goals of teaching, such as adaptability and enterprise – all these may be embodied in the assessment of "core" skills. Furthermore, many of these skills and qualities are likely to be interrelated, context-specific, and difficult to demonstrate in the school situation.

Global definitions – "personal effectiveness and expertise"

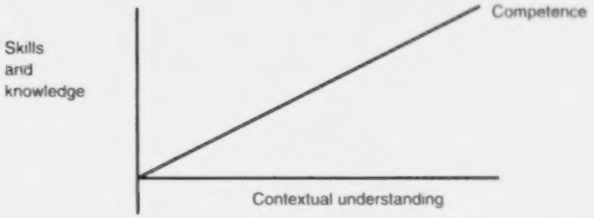
The essentially pragmatic approach to mapping the "general" outcomes of learning referred to above might usefully be related to more theoretical and research-based development work in this field to produce a more conceptually grounded but still practicable model of learning outcomes.

One attempt to articulate metacognitive skills and strategies can be found in the body of work on personal effectiveness, which assumes that metacognitive strategies centre on the ability to call upon suitable skills from a repertoire of learned behaviours as the occasion demands. This in turn relates to the ability to identify what behaviours are required for effective operation in a given situation. The "learning strategies that students adopt are powerful predictors of educational outcomes so that expertise in the selection and application of learning strategies is an important educational outcome" (Crooks, 1989). Burke (1989) terms this over-arching competence "problem-solving"; Knasel and Coates (1990) refer to it as "personal effectiveness [defined as] the ability to make things happen in a range of roles and situations ... the ability of people to make the most of their qualities and capabilities in performing appropriately across a range of roles and situations ... individuals recognising their own strengths and weaknesses" (p. 2). The relationship between the skills and qualities possessed by individuals and the ability to use them in a given situation is expressed in Figure 12.1.

The key difficulty in this model is to understand the demands of the context for overall student performance. The model might be enhanced by the addition of a conative dimension concerning the amount of effort the individual is prepared to expend in order to achieve competence in a certain domain.

Ivic (Chapter 10) highlights the Vygotskyan distinction between "manifest", "instrumental" and "structural" aspects of learning. The acquisition of a given body of content plus the application of both general and discipline-specific intellectual skills is seen as eventually leading to the longer-term development of the characteristic form of

Figure 12.1 **Model illustrating the relationship between emerging competence and contextual understanding**



Source: Stanton (1989), in Burke, p. 101.

thinking in a given subject (Elliott, 1990). Ivic's definition of expert competence which embraces all three elements plus the ability to engage in self-regulation of cognitive activity echoes the concept of personal effectiveness and many other related concepts which emphasize the ability to transfer skills and knowledge to new situations, to organise and plan work, to innovate and cope with non-routine activities, and to display interpersonal skills (Winter, 1990). Indeed they may conveniently be embraced by Aristotle's term "practical wisdom" defined in *Ethics* as at the same time a virtue, a practical interpersonal skill and a form of understanding.

As Ivic suggests, this kind of conceptualisation of learning provides a basis for structuring both curriculum and assessment procedures in terms of the development of conceptual understanding, the integration and application of learning, and the transfer of learning to new situations in a dynamic way.

New approaches to assessment

This chapter has considered the increasingly explicit call for the full range of desired learning outcomes to be reflected in assessment procedures, including those used as performance indicators. The more theoretical attempts in cognitive psychology and related disciplines to define generic learning skills as well as the pragmatic approaches to this task which are based on the perceived needs of "end users" have been taken into account. It is appropriate to turn now to the ways in which alternative approaches to assessment might help resolve some of the problems identified above. A discussion of general issues will be followed by more specific consideration of the challenges for international studies of education indicators.

Defining a framework for recording competencies

The inclusion of non-subject-specific learning outcomes in evaluation studies of student achievement appears to challenge a number of conventional assessment assumptions. Essentially, the issue of complex, idiosyncratic patterns of behaviour in non-standardized settings must be addressed. For formative purposes, the element of whether the effects of the assessment actually constitute a contribution to the learning process needs to be added to the criterion of systemic validity (Frederiksen and Collins, 1989). Within the range of summative purposes, some way of assessing and then aggregating such idiosyncratic demonstrations of general learning outcomes must be found if useful education indicators are to be derived.

The first step clearly must be the generation of a framework of desired outcomes which is appropriate for the particular purpose. If the purpose in question is the generation of internationally valid indicators of learning outcomes, it is very important that the need for such outcomes should be carefully defined. Experience with attempts to incorporate higher-level thinking skills in national assessment studies suggests that practical and technical difficulties will stand in the way of providing assessment tasks which are both comprehensive and comparable.

Experience in England, where "Standard Assessment Tasks" (SATs) are being used as part of the National Assessment Programme to provide for greater assessment validity (Whetton, 1992), suggests that detailed observation by teachers as the basis for assessing a range of desired learning outcomes poses significant practical problems not only in terms of time and classroom management but also in terms of comparability. It was found that tasks needed to be narrowly focused, since children's success rate was found to vary according to the subject contexts. Hence, the latter need to be strictly comparable. However, introducing such comparable and stylized contexts tends to reduce validity.

The message from this example is that the optimum balance between reliability and validity and between various technical, practical and educational concerns must be decided when the learning outcomes are identified. The crucial question of how far any demonstrated high-level competence is generalisable either within or across school subjects leaves two main options in terms of assessment. The first is to use descriptive reporting. In this case, the context of a given learning outcome is described and the user is left to make inferences about transfer and hence the predictive validity of the assessment. The second is to develop tight assessment specifications which list the ranges of context in which achievement should be measured. This is the approach used in both the U.K. National Assessment Programme and in its new programme of national vocational qualifications. However, in the latter the "elements of competence" do not rely on particular assessment instruments or pre-specified methods of assessment – a feature that is important for the assessment of skills in different contexts. Examples of both approaches are given in Tables 12.1 and 12.2.

For National Vocational Qualifications, the outcomes required are expressed in a Statement of Competence which consists of units of competence and, at a more detailed level, elements of competence.

But even given the clear formulation of descriptive criteria, considerable ambiguity may remain. Thus, for example, a criterion statement such as "can write a letter applying for a job", is open to a good deal of subjective interpretation on the part of the assessor unless the statement is qualified by a number of other performance criteria such as "with

Table 12.1 **An example of descriptive reporting**

Example from National Curriculum

Subject: ENGLISH

Profile component: Reading

Attainment target: the development of the ability to read, understand and respond to all types of writing, as well as the development of information-retrieval strategies for the purpose of study.

Statement of attainment (at level 9): select, retrieve, evaluate and combine information independently and with discrimination, from a comprehensive range of reference materials, making effective use of the information.

(Example: make use of techniques such as skim-reading and organisational devices such as layout, illustration and placing of visual images and text and the production of text in a number of media, drawing on these devices.)

not more than five grammatical mistakes'', ''on a word-processor'', ''in not more than 30 minutes'', and so on. A closely defined skill such as typewriting or swimming lends itself quite readily to this approach whereas general skills, almost by definition, are difficult to define in terms of unambiguous performance criteria if their essential meaning is to be preserved (Sadler, 1989).

Essentially, this is a question of defining the desirable balance between reducing ambiguity or subjectivity in assessment by the detailed specification of performance criteria and recognising the danger of a proliferation of detailed assessment requirements that is unmanageable for the teacher and produces a volume of information too large and too specific to be useful for programme evaluation.

The development of ''records of achievement'' schemes in recent years has perhaps done most to highlight this issue (Broadfoot *et al.*, 1988). Faced with a commitment to reporting the whole range of student achievement both within and outside the classroom, those responsible for developing such schemes have explored a variety of ways of recording information that is both practicable for the teacher and useful for users – students, teachers, parents, trainers, employers, and educational institutions. Two main options are the use of standard performance descriptors or prose statements that describe the activities a student has engaged in and from which, explicitly or implicitly, information about particular competencies may be deduced (Suggett, 1987).

In both cases, the problem remains one of deciding the level of detail at which assessment needs to be carried out. This is particularly difficult for general skills, where it may well be desirable to integrate information about performance generated in a number of different assessment contexts. For example, the ability to operate effectively in a group

Table 12.2 **An example of assessment specifications**

Example from National Vocational Qualification

NVQ: Building Societies: Level II

UNIT

1. Provide information and advice and promote products and services to customers.

ELEMENT

1.1 Inform customers about products and services on request.

PERFORMANCE CRITERIA:

a) features, advantages and benefits of services sufficient to the customer's request are described clearly and accurately;

b) example calculations are correct;

c) appropriate information is accessed from available resources (including Viewdata);

d) information requests outside the responsibility of the job-holder are passed on to an appropriate authority promptly and accurately;

e) customers are acknowledged promptly and treated politely;

f) customers are treated in a manner which promotes goodwill.

RANGE OF VARIABLES TO WHICH THE ELEMENT APPLIES:

Investment – instant access, higher rate notice account, regular savings.

Lending – mortgages, further advances, personal secured loans, unsecured loans, credit cards.

Insurance – property, personal, travel.

Services – foreign currency, traveller's cheques, credit card, share dealing.

Customers – minors, teenagers, 16+, middle-aged, pensioners, professional contacts, companies, non-resident groups.

might need to be assessed in several different subject areas such as technology, sport and English if a confident judgement about such a "general skill" is to be made. The problem of aggregation is therefore essentially one of retaining the validity of the initial assessment whilst losing some of the detail.

A feasible approach to aggregation involves the use of pre-specified weights for particular assessment targets and their subsequent reporting in numerical terms (as in the U.K. National Assessment Programme). The records of achievement tradition differs from this approach because it is committed to descriptive, idiosyncratic statements of achievement. Such achievements might be expressed in a range of different ways. One option would be to identify which subject goals embody a generic as well as a subject-specific competency and to record these to the level of attainment.

By the same token, it may also be possible to generate skill descriptions which are specifically concerned with expressing "general skills". A critical issue in this case would be to decide how precisely the assessment framework for recording achievement would have to be laid down in advance. The advantage of leaving the framework open is that it can then accommodate a more personalised, idiosyncratic description of individual student achievement. If the purpose is to monitor learning outcomes on a national or international basis, however, then careful pre-specification of the assessment criteria is essential.

The foregoing discussion concentrates on the way in which the diagnostic and formative recording of progress in higher-level skills might be summated for more general recording and reporting purposes. However, just as the concept of "progression" is now crucial to the design of the academic curriculum, some notion of "progression" is also important in the design of assessment systems. In practice, however, the construction of hierarchies based on an unambiguous differentiation of levels of attainment has proved to be difficult.

There are many reasons for this. For example, students may not perform consistently. Some may achieve a particular level but not be competent in the same skill at lower levels (which may suggest that the hierarchy is not unidimensional). The difficulty of constructing statements that express unambiguous hierarchies of attainment are well illustrated by the fruitless search for "grade" criteria in the English examination for the general certificate of secondary education (GCSE). These "grade-related" criteria (subsequently referred to as "grade-criteria") were to be based on national criteria rather than on any particular interpretation of these criteria in a syllabus. In practice, it was found that aggregation of the putative criteria would be a major problem, as any given grade in a subject might represent different patterns of attainment of individual criteria so that the underlying profile of strengths and weaknesses across the grades might bear little resemblance to the ultimate "average" or "composite" grade awarded (Scottish Council for Research on Education, 1977).

It largely remains to be seen whether a hierarchical model – such as the one that underpins the design of national assessments in the United Kingdom – will prove any more workable in practice. The design is based on the work associated with the development of graded assessments in many subjects. The difficulty is that, whereas the detailed criteria of success are laid down at each level, they are not organised into a map of hierarchical targets to be assessed individually and sequentially. Related problems are generating descriptive statements of attainment that do not embody multiple elements of performance and specifying in advance the criteria to be attained. These and other difficulties in generating truly hierarchical, unambiguous and yet practically useful criteria of attainment suggest that the attempt to define such criteria on an international scale and in relation to higher-order skills and general competencies is likely to prove difficult.

The alternative approach would be to generate a framework of recording and reporting which recognises and clarifies these broad competencies and which lays down enough of a structure to encourage the setting of progressively more challenging individual targets within a broad general framework. This could be provided for through some kind of descriptive matrix that attempts to define the competencies in question, the factors influencing progression within them, such as complexity, confidence, breadth of experience, transferability and examples of the different levels. This is arguably just as important as assessment's and recording's more traditional preoccupation with quantitative differentiation in terms of the level of performance achieved.

Techniques of assessment

In reviewing the progress that has been made in devising assessment techniques to support the development of a wide range of competencies, it is useful to consider some of the techniques for collecting evidence on which to base an assessment. Given that the goal is to represent the full range of learning outcomes within and beyond specific learning programmes, it is not surprising to find a general emphasis on broadening the context of assessment to include not only formal classroom settings but also a range of different, and perhaps novel, situations in which the application of knowledge can be demonstrated. Typically, these more ''descriptive'' or ''authentic'' approaches to assessment are likely to complement or replace more traditional tests with some of the following:

- Written tests of a more open-ended kind which might include one or more of the following:
  - short answer questions;
  - longer structured questions;
  - free-response questions;
  - data-response questions.
- Oral/aural assessment, especially in cases where the development of oral skills is an explicit curriculum objective as in, for example, English and modern languages, but it may also figure as part of coursework assessment in other subjects.
- Practical tests – such as the continuous assessment through observation by teachers.
- Projects, such as a single, extended piece of work unique to the candidate (*e.g.* a case-study, a practical investigation, or the construction of an artefact).
- Observation (*e.g.* cumulative check list record; student self-assessment against criteria; peer assessment; student-teacher negotiated assessment).

Implementing performance assessment

In seeking to define the characteristics of systemically valid assessments, Frederiksen and Collins (1989) identify as crucial the directness of cognitive assessment and the degree of subjectivity or judgement required in assigning a score to represent the cognitive skill. This emphasizes the need to assess skills through an interpretation of direct performance. Only in this way, argue Frederiksen and Collins, can the tendency of standardized testing to concentrate on assessing partial, low-level skills be avoided, since the state of the art does not generally allow for objective tests for directly measuring higher-order thinking skills, problem-solving strategies and metacognitive abilities (p. 29).

The value of "direct" tests is well demonstrated, according to Frederiksen and Collins, by the National Assessment of Educational Progress (NAEP) primary trait system for scoring writing tasks, in which known traits are identified as being important for successful achievement of a particular assignment. At least three further advantages follow from this approach. First, with suitable training, assessors can produce correlation scores that approach quite closely those achieved with standardized scoring methods. Second, the creation of a library of exemplary training materials can help make teachers more aware of the characteristics that would constitute good performance and therefore help to clarify learning goals. Third, students also can acquire greater clarity about the goals they are aiming at.

In practice, such an assessment approach would require, first, the provision of a representative range of "ecologically valid" tasks – some compulsory, some elective – allowing assessment of both general and idiosyncratic learning outcomes of a more creative kind. The knowledge and skills involved would need to be broken down into sub-processes, each of which would be defined by a small number of "primary traits" covering both learning processes and products. Exemplar materials and training for the assessors would be a necessary accompaniment.

Expert-performance analysis

Baker *et al.* (1990) have developed a novel approach to measuring knowledge acquisition, deep understanding and problem-solving in the context of learning history, which they feel combines both the rigour necessary to ensure confidence in the results and desirable feedback for the process of instruction and learning. Initially, they sought to use the *post hoc* scoring approach used in NAEP's primary trait assessment (National Assessment of Educational Progress, 1985, 1987). Their approach involves a trade-off between specificity in the scoring criteria used and the difficulties of combining unique assessments for more general ratings. To provide useful guidance for future instruction, it is necessary to find the optimum balance of generality in criteria so that they reflect both long-term and basic learning goals and those specific to a particular subject.

Their initial approach to this problem was to develop both general and specific criteria of quality from assessments carried out by markers – both specialists and non-specialists in the field. However, this method failed to discriminate adequately between general and specific learning outcomes. This prompted them to use a more learning-centred approach in which an analysis of the behaviour of experts in solving a problem became the basis of the assessment framework for novices. In this case, the "expert" performance criteria were found to be the use of prior knowledge and the use of text references and a real problem to direct the answer, plus an explicit effort to show interrelationships. These criteria proved useful as a basis for assessing the competence of learners both in a subject area and in relation to more general learning agendas.

Computer-based approaches

A rather different approach to developing components of expert performance relates to computerised learning programmes based on an explicit theory of learning. As learning proceeds, the computer gears each instructional step to the present needs of the learner as defined by that model (Wenger, 1987). The problem is that this approach does not lend itself to reporting and cannot normally take account of the student's initial state beyond what is relevant to the programme in question. Although it is clear that useful diagnostic information can potentially be generated by such approaches, they do not appear to lend themselves to broad purpose of systemic evaluation.

Chronometric analyses

In the same vein, Siegler (1989) explores the potential of chronometric analyses, in which solution time patterns are used to infer mental processes. The danger here is that a consistent outcome pattern may be used to infer common learning strategies in a way that is quite unjustified. Siegler argues the need for empirical documentation of learners' various problem-solving strategies, using verbal and observational techniques as a necessary first step. This approach might be modified to provide a more general account of the learning skills of a particular student.

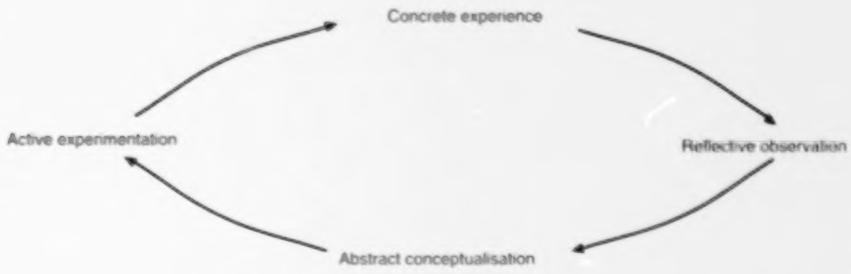
Students' self-reports

Assessment of critical thinking – in one form or another, a very general learning goal – highlights many of the generic problems associated with assessment of skill transfer. Not only is it difficult to determine how far a skill is dependent on subject-specific knowledge, it is not certain that its application in different learning domains really constitutes the same skill. Furthermore, as Norris (1989) argues, the successful demonstration of a skill involves both the ability and the disposition to use it. This latter condition points to the potential influence of the learner's pre-existing conative state, not least because it may help to determine the assumptions that students make when engaging in critical thinking. Once again students' verbal reports of their thinking processes in relation to questions posed are likely to be more illuminating for assessment purposes than the answers themselves. Norris also suggests that conventional multiple-choice tests might be used to judge whether a student can make credibility judgements. This might help to explain why it is that, in the United States at least, students appear to be able to locate information but not to use or assess it (National Assessment of Educational Progress, 1985). Norris mentions that to be adequate, such tests would need to be complemented both by the kind of process study described above and by further theoretical exploration of the nature of "critical thinking" itself.

Teach-back procedures

A rather different approach emphasizes the use of systematic interviewing or recording of student cognitive structuring (Naveh-Benjamin *et al.*, 1986). This may be linked to the use of cognitive "teach-back" procedures, in which students demonstrate their competence by teaching the same material to someone else. This performance is divided up into appropriate components and judged according to criteria such as correctness and completeness of ideas, personalisation and reorganisation of knowledge. Whilst time-consuming, this technique is flexible and has many possible applications. The rationale for this approach echoes thinking about "the reflective practitioner" in which "reflection in action" supports a spiral of increasingly sophisticated understanding ranging from the practice itself, a concrete description of that practice and, finally, reflection on the description of practice (Schon, 1987, p. 14). This learning cycle is shown in Figure 12.2.

Figure 12.2 **The experiential learning cycle**



Source: Gibbs (1988), p. 11

Assessment of problem-solving

The use of computer simulations might also be used for performance assessment. Although they may lack a measure of authenticity and are expensive to develop, they are cheap to run and score, have a high-level of reliability, and provide a full record of problem-solving activity for *post hoc* diagnostic purposes. Meyers *et al.* (1985) also emphasize generating diagnostic information through detailed, intensive observations which yield quantitative data in terms of frequency of a given performance and qualitative data in terms of its characteristics, and they stress its importance in tracing the interaction between the learner, the task and the environment in process assessment.

Another example is the paper simulation of patient management problems being evaluated by the Dutch Central Institute for Test Development (CITO) (Izard, 1990). The student is provided with a realistic case description with a given context and task, and must gather information, make a diagnosis and suggest a course of action. The responses are then scored (in future this will be done by computer to reduce the time involved).

A third approach to the assessment of problem-solving, which owes much to the British Assessment of Performance Unit (APU), is the one incorporated in early prototypes of English Standard Assessment Tasks (SATs), as the following example illustrates: The children were to use a Lego car on a ramp. They had to find out how far the car would travel from different places on the ramp, then raise the ramp and predict what would happen. They then had to choose units and decide how to measure and record their findings. They then recorded them. This activity was principally designed to assess science but also gave rise to assessments in mathematics (measuring the distances, recording information) and in English (speaking and listening, writing). The children worked in small groups, so that the assessments took place in real and familiar situations, thus enhancing their validity (Whetton, 1992).

Reference tests

A more long-term approach involves generating a course map of content and goals involving elements of knowledge and of skill development and application. The key points of transition, empirically or expertly predetermined, are the subject of "reference" tests designed to locate progress for the learner and the teacher. This approach has already been much used to create "graded" test schemes. If each reference point were also the subject of a cognitive task analysis, this would help generate information about more general learning patterns, as well as provide a specific learning trajectory in a given subject. Group profiles of learners' strengths and weaknesses at the course or multi-course level (Snow, 1989, p. 13) could be generated.

A more narrowly traditional approach to assessment – but one which nevertheless seeks to reflect the integrity of the theory of student learning which informs it, by distinguishing high-level from low-level cognitive processes – is based on the "SOLO" taxonomy (Biggs and Collis, 1982). This taxonomy identifies four stages in learning. The first is "pre-structural", in which the task itself is not approached in an appropriate way. Next come "uni- or multi-structural" levels, in which several aspects of the task are picked up serially. Then comes the "relational", in which several aspects are integrated into a coherent whole. Finally, there is the "extended abstract", in which the coherent whole is generalised to a higher level of abstraction. Biggs *et al.* (1989) provide an illustration of the model in practice in mathematics.

With regard to the more affective elements of learning, novel use can be made of a computerised dictionary of self-descriptive adjectives, using a system of weights that reflect expert scalings of each adjective on each of several personality dimensions, and thus generating personality profiles in a way that is arguably more valid than using conventional questionnaires. There is now considerable testimony to the value of self-assessment for producing valid information on conative learning states and indeed for facilitating learning itself by encouraging learners to recognise their own responsibility for learning (Broadfoot, 1977; Broadfoot *et al.*, 1988; Sadler, 1989). However, there does not seem to be a ready application of such techniques at the level of national surveys.

International studies of student achievement

Two major strands of activity may be considered in this respect – international research studies of student achievement and national monitoring procedures. This section describes briefly some case-study examples of both approaches as a basis for analysing the main problems such studies face.

Assessment in the United States

Recent decades have seen a rapid growth in the use of student assessment and testing techniques to yield information about national standards of performance. The National Assessment of Educational Progress (NAEP) in the United States was one of the first such initiatives, conducting its first survey in 1969. The assessments covered reading, writing and literacy for young adults in 1984; reading, mathematics, science, computer competence, U.S. history and literature (for 17-year-olds in 1986) and reading, writing, U.S. history, citizenship, and geography in 1988. Although the assessment instruments that were used have mostly been of the multiple-choice type, "hands-on" tasks to assess higher-order skills in mathematics and science have also been developed.

NAEP has also developed a reading scale which has the dual advantage of yielding a single numerical (and therefore comparable) score and of providing for comparisons over time, although it has been heavily criticised on both counts (Gipps, 1990, p. 8). However, by using item-response theory, test items that are good discriminators at the various levels of proficiency can be used as descriptors to express numerical scores as actual behaviours (Izard, 1990).

The concern to assess higher-order thinking skills is reflected in NAEP's adoption of material from the Assessment of Performance Unit (APU) based in England. These tasks require students to classify and sort; to observe, infer and formulate hypotheses; to detect patterns in data and to interpret results. At the most complex level, students are asked to design and conduct complete experiments (NAEP, 1987). However, problems in interpreting the performances produced and economic considerations have prompted the search for the kind of computer-based alternatives referred to above.

There has been considerable pressure for NAEP to become the basis of mandatory state-by-state comparisons. This has so far not happened and the spectre of "a national achievement test representing a national curriculum in the United States" (Ferrara and Thornton, 1988) is beginning to materialise as state-by-state studies are now being carried out (see Blank, Chapter 6 in this volume).

Assessment in England and Wales

The Assessment of Performance Unit (APU) was founded in 1975 "to promote the development of methods of assessing and monitoring the achievement of children at school and to seek to identify the incidence of underachievement". It is being replaced by a mandatory programme of national assessment tied to a mandatory national curriculum. Initially modelled on the American methodology developed for the National Assessment of Educational Progress (NAEP) during the early and mid-1980s, in practice the APU turned into a mechanism for professional development, which helped teachers to understand in more detail pupils' progress in the various subjects, to develop new insights into curriculum objectives, and to begin to apply novel assessment techniques themselves.

The first British Assessment of Performance Unit surveys took place in 1978 in mathematics, in 1979 in English language, in 1980 in science and, later, in modern languages and design for technology. Although these surveys were originally intended to monitor all aspects of learning, including aesthetic, physical, personal and social development, these areas proved impossible to conceptualise in a way that was generally acceptable for the purposes of national monitoring, and hence were abandoned. It has proved

BEST COPY AVAILABLE

impossible to fulfil the original remit of the APU concerned with identifying "underachievement" because in practice "underachievement" could not be defined. Furthermore, the technical impossibility of measuring changes in standards over time has become apparent. In practice, such surveys have therefore been reduced to using a subset of common items to gauge changes over time.

Although the Assessment of Performance Unit pioneered the development of many new assessment techniques, in oral and practical skills testing, for example, which have subsequently been incorporated into public examinations and the new national assessments, its lack of linkage to a specific curriculum and its failure to produce for each child a single comparable score equivalent to NAEP's "plausible value" have arguably led to its being replaced by a monitoring system that provides quite explicitly for comparison between students and institutions as well as offering national data on standards.

Under the English National Assessment arrangements students will follow, in each subject, a curriculum hierarchically structured into ten levels, each of which embodies a number of attainment targets – 14 in mathematics, 17 in science and six in English. Students will be assessed for all targets at the appropriate level by their teachers. In addition externally devised Standard Assessment Tasks will be administered by teachers at ages 7, 11 and 14. The results will be aggregated by teacher and school and then published.

Assessment in Australia, France and New Zealand

France follows the U.S. and British models of national monitoring by conducting large-scale surveys to gauge overall levels of performance. For example, a 1989 survey of 11-year-olds covered mathematics, reading and writing, in response to national concerns about literacy rates. Assessment techniques were of a fairly conventional kind using multiple-choice and open-ended tests. What was unusual was the emphasis placed on providing information to parents and teachers about individual students as a basis for remediation and in-service training, instead of on stimulating competition between teachers and institutions as in the United States and Great Britain.

"It was asserted that there would be no comparison between classrooms, schools and regions because the first goal of this assessment was to help teachers to improve the results of their own pupils." (Le Guen and Lacronique, 1990, p. 2)

Other countries, such as Australia, embarked on surveys similar to those in France, but abandoned them when they found they had little impact and little to say about standards. In some Australian states, attempts are now being made to introduce more explicit auditing and performance monitoring systems in the context of more devolved administration (Dimmock, 1990). Some countries, like New Zealand, are on the verge of initiating national monitoring (Bell and Dawson, 1991). The desire to engage in large-scale assessment programmes is clearly evident and is only matched by the political and technical difficulties inherent in such a project. These political and technical difficulties are arguably even more in evidence in international research studies of achievement, which embrace a range of subjects and involve a large number of countries all using different curricula.

International surveys of student achievement

The International Association for the Evaluation of Educational Achievement (IEA) began with the First International Mathematics Study in 1962. Surveys of a further six subjects – reading, literature, science, French as a second language, English as a second language, and civic education – followed and were published in 1976. Since that time, IEA has launched additional studies, including surveys of mathematics and science, written composition and reading literacy, pre-primary children, computers in education, teachers and learning, social values and morality, and classroom environment.

It is readily apparent from this list that the surveys are ambitious and go well beyond the assessment of pupil performance *per se*. In so doing, they illustrate the importance of indicators of student achievement that encompass both higher-order learning outcomes and the context for learning, but they also reveal the problems involved. Indeed, although the IEA studies were initially concerned only with surveying student outcomes, it gradually became apparent that input and process variables must also be included in order to make sense of the survey data. Ironically perhaps, the search for indicators has typically started by giving most attention to input variables, and the realisation that indicators must be addressed as part of a total framework that also includes process and output variables has come later. The increasing sophistication of the conceptual models of indicators being used has necessarily increased some of the technical difficulties in the collection and interpretation of the chosen data.

Problems in conducting international achievement studies

Many problems are associated with international studies of student achievement. These arise whatever the approach used and the outcomes being assessed. The main problems are technical (how to measure accurately the achievements in question) and political (whether the surveys provide data that are useful for policy-making purposes and the potential repercussions of such testing on the educative process itself).

Technical problems

Broadfoot (1978) and Hamilton (1977) address the understandable desire on the part of educational decision-makers not only for information about prevailing standards in their own system and how these compare with those achieved elsewhere, but also for information on the likely results of various alternative actions in policy and expenditure. In practice, however, it has proved very difficult to provide such diagnostic insights at a systemic level.

Comber and Keeves (1973) suggested that although their survey of science education in 19 countries gave an unrivalled picture of science achievement, it proved disappointing as a means of evaluating the relative efficacy of different teaching conditions and methods. To do this, they suggested, would require new ways of collecting data – a point echoed more recently by Silvester (1990) who stresses the inability of any large-scale monitoring to provide explanatory data about the quality of education provided, including quality of teaching, resources and curricular structure. As Odden (1990) suggests, "monitoring outcomes alone does not provide enough information to determine why changes in outcome occur over time" (p. 24). Furthermore, any inferences based on school-level aggregates may be prone to assuming that all individuals are equally affected, even though in reality they are not. Attempts to allow for this statistically in the population as a whole, as in the NAEP (1987) study's use of "conditioning variables" such as gender and ethnicity, lead to a substantial underweighting of school variance.

Other technical problems noted by Lapointe (1990) include the following:

- producing common tests across schools and school districts where in any but the basic skills of reading, writing and arithmetic there are unlikely to be common curricular objectives;
- ensuring validity in the test, especially when higher-order skills rather than knowledge are the focus for assessment;
- potential cultural bias embodied in the context and vocabulary of the assessments;
- the difficulties of comparison between areas where different educational priorities are pursued, different resources are provided, and different attitudes and abilities to learn prevail in the population;
- technical and psychometric issues relating to sampling, standardized administration procedures and monitoring of testing procedures.

The IEA Classroom Environment Study (Anderson *et al.*, 1989) provides an example of an explicit attempt to use a combination of qualitative and quantitative data-gathering techniques across eight countries to examine in detail those instructional variables that affect successful learning. The study was based on intensive classroom observation coupled with a pre-test and a post-test in each country. But, despite the sophistication of the data-gathering techniques used and the analytic frameworks applied, the study tells little that is new. Again, part of the reason would appear to be technical – such as the need to control for initial differences in student ability and achievement, the need to use conceptually sound and measurable variables, and the need to control "opportunity to learn" as a key variable. The need to consider the utility of various structured and unstructured approaches to collecting classroom data relates in turn to the trade-off between comprehensiveness and reliability in the data generated.

Part of the explanation for the limitations of international studies, however, would appear to be conceptual. Thus, although it was possible to identify a number of specific relationships between teacher behaviour and learning outcomes in the various countries, the assumptions underpinning the process-product analytic paradigm employed, which assumed a direct relationship between teacher behaviour and student learning, proved to be flawed. This is partly due to the familiar technical problem within IEA of using correlational and regression methods to analyse the data, despite the fact that the median contribution of "schooling" variables to overall outcomes was only 15 per cent. It is also a reflection of a more fundamental flaw in the underlying theoretical conception of the study, Anderson recognises, citing Dunkin and Biddle's (1974) work on teaching:

> "Perhaps the greatest single flaw in much of the research we have reviewed is the persistent assumption that appears to underlie much of it that teaching can somehow be reduced to a scalar value that can be indicated by a frequency of occurrence for some teaching behaviour. We suspect ... that this simply is not true." (p. 353)

The above criticism is only one reflection of the endemic tendency in assessment to use often invalid, indirect measures of the variable in question. Resnick and Resnick (1989) provide another illustration, arguing that complex competencies such as "thinking" cannot be assessed in the decomposed and decontextualised approach assumed in standardized testing, since the organic whole of such a task is, like teaching, more than the sum of its various elements. Examining tests used in mandated state testing programmes of educational quality in the United States, Resnick and Resnick argue that the short and superficial questions used, for example, in reading comprehension cannot test, or give students opportunity to demonstrate, higher-order thinking skills. Although progress in introducing more authentic assessment measures to address this issue is clearly being made, it is apparent that in order to avoid excessive cost, either a political compromise will be needed to accept relatively light or infrequent sampling, or new surrogate assessment approaches will have to be developed to reduce the expense of employing trained observers.

The limitations of existing assessment techniques coupled with political expediency can also lead to employing undesirable conceptions of achievement, as McLean (1990) demonstrates in his analysis of the IEA Reading Literacy survey. There, mainstream scholarship in the field was set aside in the interest of political and practical considerations (p. 76). Thus, it seems inevitable that, as McLean (1989) suggests, authenticity in assessment will be inversely related to traditional estimates of measurement quality. Indeed, as far as the more complex learning outcomes are concerned, it is still hard to see how they might be incorporated into large-scale surveys, given the prevailing assessment preoccupations and the slow development of acceptable alternative techniques.

Problems of application

The second major problem with national and international surveys of learning outcomes centres on the use made of the results of such studies. Wedman *et al.* (1990), reviewing evidence from national assessment programmes in the United States, the United Kingdom and Australia, concluded that lack of use of the information obtained was a common problem. In Australia, for example, the reaction of political bodies, media, teachers, teacher unions, parents and community groups to the 1980 national testing results was negligible. The same picture is painted by Gipps and Goldstein (1983) in the United Kingdom. However, in many countries this situation is changing.

All these relatively well-known criticisms of national and international assessment programmes centre on the need for explanation. Indicators of performance need to have an "interpretive frame of reference" (Wedman *et al.*, 1990) in which indicators are defined as "individual or composite statistic(s) that relate to a basic construct in education and are useful in a policy context" (Shavelson *et al.*, 1989, p. 5).

Before concluding this section it is appropriate to consider how far existing surveys of this kind can shed light on the quest for indicators of higher-order skills and competencies. Some possibilities are immediately apparent. These include:

- complementing test results with expert reports, as in France, or with observations of classroom performance, as is done by the NAEP in the United States;
- using survey results to calibrate more authentic "hands-on" testing;
- designing Standard Assessment Tasks, as in England, which explicitly map on to curriculum objectives of all kinds;
- incorporating a variety of assessment results in an explicitly descriptive "record of achievement" or in the form of direct evidence as a portfolio;
- putting more emphasis on assessment by teachers (and the associated need to provide them with training in assessment).

Underlying all of the above is the need for precise definitions of subject domains; for desired learning outcomes; for a trade-off between different assessment techniques and how their various strengths and weaknesses may be most fruitfully combined for a given purpose; of the putative audience for the assessment and that audience's primary need, recognising that differing assessment purposes – such as diagnosis and accountability – may well be mutually incompatible; and of the overall political context which will influence what can be and is done.

Conclusion

Thus, it may truly be said that "we want educational indicator systems to accomplish objectives beyond our current knowledge base. We do not know how all the critical components of the education system work to produce outcomes, yet we are engaged in the process of developing indicator systems that must be designed to do just that" (Odden, 1990, p. 24).

The current lack of techniques to measure students' performance in higher-order cognitive skills and general learning outcomes as an important individual indicator is well recognised by those involved. If this lack continues, the consequences are also well recognised, and many of them have been reviewed here. There is clearly an urgent need:

> "to encourage the inventiveness and adaptability of educational systems by developing tests that directly reflect and support the development of the aptitudes and traits they are supposed to measure." (Frederiksen and Collins, 1989, p. 28)

Such tests need to be characterised by their focus on actual performance; their comprehensive coverage of learning goals; their application of clearly defined hierarchical criteria, as the basis for both reliability and for transparency, so that those being assessed understand the criteria being applied and can assess themselves and direct their learning appropriately. They need to be able to address both domain-specific and general learning goals; both short-term and long-term desired outcomes; to combine cognitive and conative dimensions; to be usable in a range of potential performance contexts – verbal, symbolic, physical and social; and to take account of idiosyncratic, unanticipated learning as well as instructional goals. The data also need to be generated in a form that allows for aggregation across individuals and to be relevant for making policy decisions.

The international interest in the development of effective education indicators (Walberg, 1990) means there is a very real danger of invalid, shallow data and inappropriate correlations being generated in an attempt to short-cut the necessary development work that still needs to take place in designing new assessment approaches. A particular subject-based approach to the definition of learning goals and their assessment has predominated for over 100 years, in which the canons of psychometric assessment have emphasized the need for objectivity as an overriding goal. The establishment of a very different set of assessment principles among both educational professionals and the wider population is likely to take a good deal of time and effort. As with most other problems in assessment, the issues are both technical and political; they involve generating suitable techniques and rendering them acceptable for the various social purposes that assessment fulfils.

This chapter has described some of the steps that have already been taken towards this goal. "To the extent that assessment experts prove themselves able to meet the challenge of the development of a new theory of test design and validation – one that emphasizes individual learning rather than individual differences" (Baker *et al.*, 1990, p. 32), they will have done far more than provide some elements of a more valid system of education indicators, important as this is. The "backwash" effect will ensure that education systems around the world will begin to direct their efforts towards generating those vital intellectual, social and practical competencies that for so long have been the subject of largely empty rhetoric on the agenda.

Bibliography

ANDERSON, L.W., RYAN, D.W. and SHAPIRO, B.J. (eds.) (1989), *The IEA Classroom Environment Study*, Pergamon Press, Oxford.

ASSESSMENT OF PERFORMANCE UNIT (APU) (1984), "What it is and how it works", Department of Education and Science, London.

BAKER, E.L., FREEMAN, M. and CLAYTON, S. (1990), "Cognitive Assessment of Subject Matter: Understanding the Marriage of Psychological Theory and Educational Policy in Achievement Testing", paper read at the Annual Conference of the American Educational Research Association, Boston, MA.

BELL, C. and DAWSON, D. (1991), "Monitoring System for Schools Standards Planned", *The Dominican Sunday Times*, 10 February, New Zealand.

BIGGS, J.B. and COLLIS, K. (1982), *Evaluating the Quality of Learning*, Academic Press, New York.

BIGGS, J.B., HOLBROOK, J.B., KI, W.W., LAM, R.Y.H., LI, W.O., PONG, W.Y. and STIMPSON, P.G. (1989), "An Objective Format for Evaluating the Quality of Learning in Various Secondary Subjects: A Symposium", Hong Kong Educational Research Association, Sixth Annual Conference, Hong Kong.

BLOOM, B.S., ENGLEHART, M.D., FURST, E.J., HILL, W.H. and KRATHWOHL, D.R. (1956), *Taxonomy of Educational Objectives: Handbook 1, Cognitive Domain*, Longman, London.

BROADFOOT, P. (1977), "Communication in the Classroom: The Role of Assessment in Motivation", *Educational Review*, Vol. 31 (3), pp. 3-11.

BROADFOOT, P. (1978), "The IEA Studies in Evaluation: Some Views on the Six Subject Surveys", *Prospects*, Vol. 8 (1), pp. 121-126.

BROADFOOT, P., JAMES, M., McMEEKING, S., NUTTALL, D. and STIERER, B. (1988), *Records of Achievement: Report of the National Evaluation of Pilot Schemes*, Department of Education and Science, London.

BURKE, J.W. (ed.) (1989), *Competency-based Education and Training*, Falmer Press, Falmer, Sussex.

COLE, N. (1990), "Conceptions of Educational Achievement", *Educational Researcher*, Vol. 19 (3), pp. 2-7.

COMBER, L.C. and KEEVES, J.P. (1973), *Science Education in Nineteen Countries: An Empirical Study*, John Wiley, New York.

CRONBACH, L.V. (1984), *Essentials of Psychological Testing* (4th edition), Harper and Row, New York.

CROOKS, T. (1989), "The Impact of Classroom Evaluation Practices on Students", *Review of Educational Research*, Vol. 58 (4), pp. 438-481.

DEPARTMENT OF EDUCATION AND SCIENCE (1989), *Report of the Records of Achievement*, National Steering Committee, Department of Education and Science and the Welsh Office, London.

DIMMOCK, C. (1990), "Managing for Quality, Accountability in Western Australian Education", *Education Review*, Vol. 42 (2), pp. 197-206.

DUNKIN, M. and BIDDLE, B.J. (1974), *The Study of Teaching*, Holt, Rinehart and Winston, New York.

ELLIOTT, J. (1990), "Competency-based Training and the Education of the Professions: Is a Happy Marriage Possible?", paper presented at the Annual Conference of the British Educational Research Association, London.

FERRARA, S.F. and THORNTON, S.J. (1988), "Using NAEP for Interstate Comparisons: The Beginnings of a National Achievement Test and National Curriculum", *Educational Evaluation and Policy Analysis*, Vol. 10 (3), pp. 200-212.

FREDERIKSEN, J.R. and COLLINS, A. (1989), "A Systems Approach to Educational Testing", *Educational Researcher*, Vol. 18 (9), pp. 27-32.

GAGNE, P.M., BIGGS, L.J. and WAGER, W.W. (1988), *Principles of Instructional Design*, Holt, Rinehart and Winston, New York.

GIBBS, G. (1988), *Learning by Doing*, Further Education Unit (FEU), London.

GIPPS, C. (1990), "National Assessment: A Comparison of English and American Trends", in Broadfoot, P., Murphy, R. and Torrance, H. (eds.), *Changing Educational Assessment*, Routledge, London.

GIPPS, C. and GOLDSTEIN, H. (1983), *Monitoring Children – An Evaluation of the Assessment of Performance Unit*, Heineman, London.

HAERTEL, E. (1985), "Construct Validity and Criterion-referenced Testing", *Review of Educational Research*, Vol. 55 (1), pp. 23-46.

HAERTEL, E. (1986), "Choosing and Using Classroom Tests: Teachers' Perspectives on Assessment", paper read at the Annual Conference of the American Educational Research Association, San Francisco, CA.

HAMILTON, D. (1977), "The IEA Studies in Evaluation: Some Views on Volumes 7-9", in *Prospects*, Vol. 7 (3), pp. 210-217.

HIRST, P. (1974), *Knowledge and the Curriculum*, Routledge and Kegan Paul, London.

HER MAJESTY'S INSPECTORATE (HMI) (1977), *Curriculum 11-16 Working Papers: A Contribution to Current Debate*, Her Majesty's Stationery Office, London.

INGENKAMP, K. (1977), *Educational Assessment*, NFER/Nelson, Windsor.

IZARD, J. (1990), "Assessing Learning Achievement", paper prepared for the UNESCO survey of international trends in the assessment of learning achievement, UNESCO, Paris (mimeo).

KNASEL, E. and COATES, P. (1990), *Personal Effectiveness and Further Education*, Training Agency, London.

LAPOINTE, A.E. (1990), "Testing in the USA: Political and Social Aspects", in Schreiber, W.H. and Ingenkamp, K. (eds.), *International Developments in Large-Scale Assessment: Proceedings of the Symposium on Large-Scale Assessment in an International Perspective*, 16-18 June 1988, Diesheim, Germany, and NFER/Nelson, Windsor.

LEE-SMITH, M. (1990), 'Meaning of Test Preparation", paper read at the Annual Conference of the American Educational Research Association, Boston, MA.

LE GUEN, M. and LACRONIQUE, C. (1990), "Evaluation CE2-6ème: A Survey Report of Assessment Procedures in France on Mathematics, Reading and Writing", paper prepared for the IAEA Conference, Maastricht.

McLEAN, L.D. (1989), "Time to Replace the Classroom Test with Authentic Measurement", *The Alberta Journal of Educational Research*, Vol. 36 (1), pp. 78-84.

McLEAN, L.D. (1990), "Possibilities and Limitations in Cross-national Comparisons of Educational Achievement", in Broadfoot, P., Murphy, R. and Torrance, H. (eds.), *Changing Educational Assessment*, Routledge, London.

MEYERS, J., PFEFFER, J. and ERLBAUM, V. (1985), "Process Assessment: A Model for Broadening Assessment", *Journal of Special Education*, Vol. 19 (1), pp. 73-89.

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP) (1985), *The Reading Report Card: Progress toward Excellence in our Schools, Trends in Reading over Four National Assessments: 1971-1984*, Educational Testing Service, Princeton, NJ.

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP) (1987), *Manual: Learning by Doing*, Educational Testing Service, Princeton, NJ.

NATIONAL CURRICULUM COUNCIL (1990), "Core Skills 16-19 – A Response to the Secretary of State", National Curriculum Council of England and Wales, London (mimeo).

NAVEH-BENJAMIN, N., McKEACHIE, W.J., LIN, Y.G. and TUCKER, D.G. (1986), "Inferring Students' Cognitive Structures and Their Development Using the 'Ordered Tree Technique'", *Journal of Educational Psychology*, Vol. 78 (3), pp. 130-140.

NICKERSON, R.S. (1989), "New Directions in Educational Assessment", *Educational Researcher*, Vol. 18 (9), pp. 3-7.

NORRIS, S.P. (1989), "Can We Test Validly for Critical Thinking?", *Educational Researcher*, Vol. 18 (9), pp. 16-21.

OATES, T. (1990), "A Review of Core-competencies Research", Further Education Staff College, Coombe Lodge, Somerset (mimeo).

ODDEN, A. (1990), "Educational Indicators in the United States: The Need for Analysis", *Educational Researcher*, Vol. 19 (4), pp. 24-29.

RAVEN, J. (1989), "Questionable Assumptions in Test Construction", *Bulletin of the International Test Commission*, No. 28 and 29, London.

RESNICK, L.B. and RESNICK, D.P. (1989), "Assessing the Thinking Curriculum: New Tools for Educational Reform", in Gifford, B.R. and O'Connor, M.C. (eds.), *Future Assessment: Changing Views of Aptitude, Achievement and Instruction*, Kluwer, Boston.

SADLER, R. (1989), "Formative Assessment and the Design of Instructional Systems", *Instructional Science*, Vol. 18, pp. 119-144.

SCHON, D. (1987), *Educating the Reflective Practitioner*, Jossey-Bass, San Francisco, CA.

SCOTTISH COUNCIL FOR RESEARCH AND EDUCATION (1977), *Pupils in Profile*, Hodder and Stoughton, London.

SHAVELSON, R.J., McDONNELL, L.M. and OAKES, J. (eds.) (1989), *Indicators for Monitoring Mathematics and Science Education*, The Rand Corporation, Santa Monica, CA.

SHEPARD, L.A. (1990), "Psychometricians' Beliefs about Learning", paper presented at the Annual Conference of the American Educational Research Association, Boston, MA.

SIEGLER, R.S. (1989), "Strategy, Diversity and Cognitive Assessment", *Educational Researcher*, Vol. 18 (9), pp. 15-20.

SILVESTER, P.J. (1990), "The Assessment of Performance Unit: From Psychometric Survey to Diagnostic Analysis", in Schreiber, W.H. and Ingenkamp, K. (eds.), *International Developments in Large-Scale Assessment: Proceedings of the Symposium on Large-Scale Assessment in an International Perspective*, 16-18 June 1988, Diesheim, Germany, and NFER/Nelson, Windsor.

SNOW, R.E. (1989), "Towards Assessment of Cognitive and Conative Structures in Learning", *Educational Researcher*, Vol. 18 (9), pp. 8-14.

STANTON, G. (1989), "Curriculum Implications", in Burke, J.W. (ed.), *Competency-based Education and Training*, Falmer Press, Falmer, Sussex.

SUGGETT, D. (1987), *Guidelines for Descriptive Assessment*, Victorian Institute for Secondary Education (VISE), Melbourne.

WALBERG, H.J. (1990), "OECD Indicators of Educational Productivity", *Educational Researcher*, Vol. 19 (4), pp. 30-33.

WEDMAN, I., KARP, S., LYXELL, T., ANDERSSON, E. and ASTRAM, T. (1990), "Feedback of Assessment Results to Teachers, Students and Parents: The Role of Indicators", in Schreiber W.H. and Ingenkamp, K. (eds.), *International Developments in Large-Scale Assessment: Proceedings of the Symposium on Large-Scale Assessment in an International Perspective*, 16-18 June 1988, Diesheim, Germany, and NFER/Nelson, Windsor.

WENGER, E. (1987), *Artificial Intelligence and Tutoring Systems*, Kaufmann, Los Altos, CA.

WHETTON, C. (1992), "Context-based Measures for the Quality of Education: Practices and Limitations – Examples from England", in Veddar, P. (ed.), *Measuring the Quality of Education*, Swets and Zeitlinger, Amsterdam.

WINTER, R. (1990), "Beginning to Identify Post-Qualifying Professional Competencies in Social Work", Assessment Programme Working Paper, No. 1, British Educational Research Association, London.

Chapter 13

Labour Market Outcomes as Indicators of Educational Performance

by

Russell W. Rumberger
University of California, Santa Barbara, United States

Education and the economy

Throughout the industrialised world, education is viewed as the key to economic competitiveness. The major industrialised nations have come to realise that they must not only compete with one other in the global market place, but with a growing number of newly industrialised countries, such as Korea, Taiwan, Singapore, and Hong Kong, that can now produce sophisticated technological products at much lower wages. The key to survival in this new economic climate is for the industrialised nations to raise worker productivity and product quality that can justify higher wages and maintain a high standard of living. In order to do this, these nations' workforces must be more educated and better trained:

> "As the economies of developed nations move further into the post-industrial era, human capital plays an ever more important role in their progress. As the society becomes more complex, the amount of education and knowledge needed to make a productive contribution to the economy becomes greater." (Johnston and Packer, 1987, p. 116)

Of particular concern is that the use of new technologies and new forms of work organisation, which are being used to raise productivity, are greatly increasing the skill and educational demands of work. The increased use of computers, sophisticated communication systems, work teams, and other technological and organisational innovations are all claimed to require higher levels and more varied skills in the workers who are using them (Rumberger and Levin, 1989). And since such innovations are becoming more widespread in virtually every sector of the economy, the overall skill level of the workforce needs to be dramatically improved. As the importance of education for economic well-being has grown in the eyes of government, business, and the general public, so too has the dissatisfaction with the current education systems in many industrialised countries. In the United States, a litany of reports by government, business, and social leaders have decried the sorry state of the nation's education system. A variety of indicators – high school completion rates, changes in test scores over time, and particularly international comparisons of educational achievement – have shown the education system of the United States to be lacking. The concern is that without substantial improvements in education, the country will fall behind its competitors in the international market place:

> "Our trading partners have realised that their productivity will determine both their international power and standard of living. These countries have made substantial commitments to educate and train their workforces. America has, in many respects, failed to do the same." (U.S. Department of Labor, 1989, p. 1)

Of course, the United States is not alone in its concern over the state of its education system. Many other industrialised countries are also concerned whether their education systems are adequate to meet the challenge of the new international competitiveness (OECD, 1989; European Round Table of Industrialists, 1989).

The growing recognition of the importance of education in improving economic performance, and the increasing dissatisfaction with the current state of education, have led industrialised countries to undertake significant reforms of their education systems. In the United States, where primary authority for the provision of public education lies at the state level, a number of states have enacted sweeping reforms to raise funding, to restructure schools, and to improve student performance (Bacharach, 1990). Many other countries have also responded to these concerns with major reforms or proposals for reforms of their education systems (OECD, 1989; European Round Table of Industrialists, 1989).

Concomitant with these developments has been a growing interest in indicator systems to monitor educational performance (*e.g.*, Shavelson *et al.*, 1987; Murnane, 1987). As government, business, and public interest in education and international competitiveness have grown, so too has the perceived need for better and more timely information on how the education system is performing. This need is also fuelled by the increased availability of regional, national, and international data on educational performance and the computer technology to process such data. In addition, policy-makers are demanding performance information in order to show that recent reforms in education are working. In California, for example, a recent funding increase for public schools was coupled with a mandate for districts to develop broad measures of educational performance (California State Department of Education, 1989).

One area that has not received much attention in discussions of education indicator systems is the need to monitor the economic outcomes from schooling. Most monitoring systems proposed to education focus only on the direct educational outputs from schooling, such as attainment levels, educational achievement, and perhaps attitudes and values. Yet a considerable body of economic research has demonstrated convincingly that education also produces a wide variety of economic outcomes, benefiting both individuals and society at large. If policy-makers are concerned whether their nations' education systems are adequately responding to the needs of the economy, then they will need additional information on the economic outcomes that result from schooling.

The purpose of this chapter is to discuss the use of economic outcomes as measures of educational performance. The next section will discuss the rationale for focusing on economic outcomes from schooling based on existing theoretical and empirical research. The second section then presents a number of limitations and complications to the perceived relationship between educational outputs and economic outcomes. The third section presents a basic conceptual model that can be used as a basis for developing labour market indicators of educational performance. The fourth section then discusses a series of labour market indicators based on the conceptual model as well as some additional issues that must be addressed in order to develop operational measures of these indicators. The final section draws some implications for the use of this information.

Rationale for focusing on economic outcomes

There is a substantial body of both theoretical and empirical research that shows a direct and strong relationship between educational outputs and a wide variety of social, political and economic outcomes. These outcomes accrue to both individuals and society at large. For instance, individuals benefit from increased levels of schooling by receiving higher salaries, getting better jobs, and having access to additional education and training. Individual benefits also produce social benefits in the form of improved productivity and economic growth, higher tax revenues, and reduced demands for welfare and social support services. There are other benefits as well. This section will review briefly the rationale for focusing on the economic outcomes from education.

Pecuniary outcomes

Of all the benefits associated with education, the one that has been subject to the most research and enjoys the most political attention is the wage or pecuniary benefit from education. There are several reasons for this attention. First, pecuniary benefits are very meaningful to both individuals and governments, as they enable people to achieve a higher standard of living and often enjoy higher social status, while enabling society to achieve a higher overall standard of economic well-being. Second, data on pecuniary benefits – wages and salaries – are routinely collected and available, thus facilitating ready accounting and analysis of this benefit. Finally, other economic and social benefits of education either accrue directly or indirectly from pecuniary benefits, as the following discussion will point out. Thus analysing the pecuniary benefits from education can give an indication of the other benefits from education.

The most common and popular theoretical explanation for the pecuniary benefits from education is based on human capital theory. This theory was developed in the late 1950s and early 1960s by Schultz, Becker and other neo-classical economists to explain the well-known relationship between an individual's years of schooling and his or her earnings in the labour market (Schultz, 1961; Becker, 1964). The explanation is based on the tenets of neo-classical economic theory in which individuals and firms attempt to maximise their personal well-being through the operation of the market place.

Briefly, *individuals* invest in education, training and other activities that increase their human capital based on their tastes and preferences, the costs of investment (both direct costs and indirect or foregone income), and the expected economic benefits. *Firms* hire all factors of production – human capital, physical capital, and land – based on available technologies and market prices until the value of the marginal product associated with each factor equals the costs of acquiring it. *Markets* regulate supply and demand for human capital through the price mechanism based on unrestricted competition and ready access to information.

Human capital theory has provided the basis for a substantial body of empirical research on the pecuniary benefits from education. Beginning in the 1960s, economists

Table 13.1 **Relative wages[a] of workers by education level in the United States, 1963-1986**

Years	High school	Some college	College	Graduate school
All experience groups				
1963-68	10.7	16.7	31.4	13.6
1969-74	9.5	17.1	34.2	14.2
1975-80	11.0	12.6	33.8	16.9
1981-86	14.2	14.8	37.6	17.7
1-5 years of experience				
1963-68	18.8	12.7	26.7	10.9
1969-74	14.2	8.9	30.4	12.8
1975-80	17.2	8.3	22.6	12.6
1981-86	19.3	14.8	34.1	12.6

a) Relative wages represent the average wages of one schooling group relative to the average wages of the next lowest schooling group.

Source: Kevin Murphy and Finis Welch (1989), "Wage Premiums for College Graduates: Recent Growth and Possible Explanations", *Educational Researcher*, 18, 4 (May), Table 2.

have documented the strong association between education and earnings for various demographic groups (race, ethnicity, gender, age), geographic regions, and periods of times (Becker, 1964; Mincer, 1974). In some cases the technique of rate-of-return analysis is used to compare the expected economic benefits to schooling over the lifecycle to the costs of that investment (Cohn and Geske, 1990).

Table 13.1 offers an illustration. It shows the results of a study which examined the relative wages of workers in the United States with various levels of education (Murphy and Welch, 1989). The data show that young workers with one to three years of college had wages that were 8 to 15 per cent higher over this period than workers who had completed a maximum of 12 years of schooling. Workers who had completed four years of college had wages that were 23 to 34 per cent higher than workers with only a few years of college. Finally, workers with graduate school training (more than four years of college) had wages that were 11 to 13 per cent higher than those with four years of college. Relative wages have also varied over time, with college graduates entering the labour market in the 1970s having lower relative wages than workers who entered the market in the 1980s.

In addition to empirical studies documenting the individual pecuniary benefits to education, other research has demonstrated the social benefits associated with education. Rate-of-return analysis has been employed to estimate the social benefits to investment in education relative to the costs (McMahon and Geske, 1982). Using national accounting data, Denison analysed the various factors that contributed to economic growth in the United States since 1929, including the educational attainments of the American workforce:

> "Educational background decisively conditions both the types of work a person is able to perform and his proficiency in any particular occupation. A continuous upward shift in the educational background of the American labour force has upgraded the skills and versatility of labour and contributed to the rise in national income." (Denison, 1979, p. 42)

While much of the theoretical and empirical research on the pecuniary benefits to education has come from the United States, research conducted in other industrialised and developing countries has also shown large pecuniary returns to education. In general, rates of return to education are similar across countries, although they tend to be somewhat higher in developing countries where the level of educational attainment is lower (Psacharopoulos and Woodhall, 1985; Psacharopoulos, 1989).

Other economic outcomes of education

In addition to the pecuniary benefits from education, a wide variety of other economic outcomes have been attributed to education. As in the case of pecuniary returns, these benefits accrue to both individuals and society at large. The benefits are listed in Table 13.2. In some cases, there is considerable research to demonstrate the effect of education on these outcomes. In other cases, the effects have been claimed, but not substantiated with considerable empirical research.

Table 13.2 **Economic outcomes associated with education**

Individual outcomes	Social outcomes
Higher labour market earnings	Higher tax revenues
Higher non-wage remuneration	Improved economic growth
Additional education and training	Reduced economic inequality
Better employment	Better production of knowledge
More efficient labour market search activity	Reduced crime
Better child care quality	Improved political participation
Individual and family health	Reduced demand for social services
Better family decisions	Improved national health
More efficient consumer choice	Improved intergenerational mobility
Better attitudes and behaviour	

Sources: Haveman, R. H. and B. L. Wolfe (1984), "Schooling and Economic Well-Being: the Role of Nonmarket Effects", *Journal of Human Resources*, 19, 3 (Summer), pp. 377-407; Levin H. M. (1972), *The Costs to the Nation of Inadequate Education. Study prepared for the Select Committee on Equal Educational Opportunity*, U.S. Senate. Washington, D.C.: U.S. Government Printing Office; Windham, D. M. (1988), "Effectiveness Indicators in the Economic Analysis of Educational Activities", *International Journal of Educational Research*, 12, pp. 575-666.

At the individual level, several additional economic benefits from education are realised in the labour market. First, educated workers are more likely to find jobs and to keep them in periods of economic downturn than other workers (Oi, 1962). Second, more educated workers not only enjoy higher earnings than less educated workers, they also

BEST COPY AVAILABLE

receive higher non-wage remuneration in the form of better working conditions and fringe benefits (Mathios, 1989). Third, educated workers are more likely to have access to and invest in further education and training that leads to additional economic benefits over their working lives (Mincer, 1989).

In addition to the economic benefits realised in the labour market, several other social benefits from education have been identified. First, parental education is said to enhance the quality of child care and children's educational performance (Leibowitz, 1974; Hill and Stafford, 1974). Second, individual and family health are also improved (Fuchs, 1974; Grossman, 1975). Third, education is said to lead to more informed fertility decisions (Michael, 1973). Finally, it is said to enhance consumer efficiency (Michael, 1972).

At the societal level, education has a number of important economic and social benefits. It contributes to economic growth by improving the productivity of workers. Education also contributes to economic growth by improving the stock of knowledge in society which leads to improved techniques of production (Denison, 1979). The pecuniary benefits of education also lead to higher tax revenues for government.

In addition to these direct economic benefits, there are a number of social benefits that have important economic consequences. First, by raising income, increased education reduces the number of individuals and families living in poverty and thus reduces government outlays for social services (Ribich, 1968). Similarly, improved health levels associated with higher levels of schooling reduce private and public outlays for health care (Levin, 1972). Raising education can also help reduce crime rates, thereby reducing social costs of processing and incarcerating criminals.

The economic benefits from these activities can be sizeable. One recent study that analysed the social and economic benefits in the city of Los Angeles estimated that the forgone income associated with one cohort of students who failed to complete secondary school in 1986 was $3.2 billion and the social costs to local government of funding criminal services, welfare, and health attributable to this cohort was $488 million (Catterall, 1987, Table 4).

Other social outcomes that are also claimed to be associated with education include increased political participation and improved inter-generational mobility (Levin, 1972; Rumberger, 1983). Finally, education is said to play an important role in the *distribution* as well as the *level* of income in society, although it is not clear from the available evidence whether it contributes to or actually helps reduce economic inequality (Jencks *et al*, 1972; Levin, 1976; Levy, 1988).

As in the case of research on pecuniary benefits, much of the research on other economic benefits from education comes from the United States. Yet there is still a growing body of research on the impact of education on these outcomes for other industrialised and developing countries as well (Blaug, 1978; Psacharopoulos, 1987).

Difficulties in assessing the economic benefits of education

The foregoing discussion suggests there is a relatively simple and straightforward relationship between educational outcomes and a variety of economic outcomes. The strongest theoretical support for this perspective comes from human capital theory. The human capital explanation of the relationship between education and economic outcomes rests on three primary propositions:

- The primary role of formal schooling is to develop the human capital, or the knowledge and skills, of future workers.
- The labour market efficiently allocates educated workers to firms and jobs where these are required.
- The human capital of workers increases their productivity in the workplace which is then rewarded with higher earnings.

Each of these propositions can be supported with existing research. But each has also been the subject of considerable challenges by competing theories and research. The following review identifies some of those challenges.

The role of education

In human capital theory, formal schooling is the primary mechanism for developing the initial stock of knowledge and skills that entry-level workers bring into the labour market. Once they enter the labour market, workers can increase their stock of human capital by investing in formal training programmes and in informal, on-the-job training (Mincer, 1989). This additional investment in human capital formation is the primary reason that workers' earnings continue to increase during the beginning of their working lives and then decrease in later years (Cohn and Geske, 1990).

But a number of criticisms have been levied against the notion of human capital and its relationship to formal schooling. Some of these criticisms suggest that the notion of human capital formation in schooling is simplistic and incomplete. Other criticisms suggest the human capital explanation of the observed relationship between schooling and earnings is wrong. Both types of criticisms suggest that the relationship between education and labour market outcomes is more complex than the simple tenets of human capital theory suggest.

One criticism argues that human capital theory largely ignores *qualitative* differences in schooling and the impact of these differences on the formation of human capital. One form of this criticism holds that the amount of human capital developed in schools depends not simply on the amount of time that individuals spend in school, most often measured by years of schooling completed, but also on the quality of the learning that takes place which, in turn, depends on the quality of school inputs such as teachers and other resources (Behrman and Birdsall, 1983; Hanushek, 1986). Another form of this criticism argues that the human capital theory ignores important qualitative differences in the *types* of skills produced in schools. While human capital theory does acknowledge differences in *general* work skills that can be applied to a wide variety of jobs, and *specific* work skills that can be applied to a limited number of jobs or work settings (Becker, 1964), a large literature in psychology suggests that there are important independent dimensions to human abilities and skills that cover not only cognitive areas, but also physical and social areas (Gardner, 1983; Wagner and Sternberg, 1984). In fact, some critics argue that the primary role of schooling is to socialise individuals by developing a wide range of non-cognitive abilities and other traits – including proper attitudes and values such as respect for authority, proper behaviours such as punctuality, and even appropriate manners of speech and modes of dress – that will help to make them more productive workers in the large bureaucratic institutions in which most will work (Bowles and Gintis, 1976).

A more fundamental criticism of human capital theory argues that the primary function of education is not so much to *develop* human capital, but rather to *screen* individuals. In screening models, formal schooling serves primarily as a mechanism for identifying and sorting individuals who already have the abilities and other traits that are required in the workplace (Taubman and Wales, 1974). The credentials that schools award simply serve as "signals" for employers to identify more able workers and they provide workers with access to different types of jobs with different earnings, status, and promotion opportunities (Spence, 1973).

The most fundamental and radical challenge to the human capital role of schooling comes from those who view schooling as the primary mechanism for reproducing the unequal economic and social relations in the larger society. In this perspective, investment in schooling is not simply predicated on individual tastes and preferences, but is conditioned by one's social background (Bowles and Gintis, 1976; Carnoy and Levin, 1985). Moreover, large social class differences lead to large differences in schooling, such that lower-class students receive less schooling, and qualitatively different, than upper-class students in order to help prepare them for different positions in the job hierarchy (Weis, 1988).

Together, these challenges argue that schools either identify or actually produce a large number of different types of outcomes that may be valuable in the workplace. These outcomes not only include basic knowledge and cognitive skills, such as reading, writing and thinking, but social skills, appropriate attitudes, values and behaviours, and proper modes of speech and dress.

The functioning of the labour market

A second important proposition to support the human capital perspective is that the labour market effectively allocates educated workers to firms and jobs where their human capital is required. For the labour market to be the appropriate mechanism for allocating educated labour, at least three conditions must be satisfied: *i)* labour markets must be competitive; *ii)* information must be freely available; *iii)* prices must reflect relative scarcities. Each of these conditions has been questioned in the research literature.

First, a growing body of research has questioned whether labour markets are competitive. Two alternative notions have been advanced. One argues that many jobs are filled through "internal" labour markets that operate *within* organisations rather than conventional, "external" labour markets posited by neo-classical theory (Doeringer and Piore, 1971). The other argues that the external labour market is actually segmented into a number of distinct, relatively independent markets in which different types of workers (*e.g.* low versus high skilled, women versus men) compete for different types of jobs with different earnings, promotion possibilities and other characteristics (Gordon *et al.*, 1982).

Second, scholars have questioned how freely information flows to labour market participants. Information is not uniformly available to all participants, resulting in more favourable opportunities and treatment for some individuals than for others (Osterman, 1980). Also, even with relevant and timely information, there is a lag between existing labour market conditions and the time when students make decisions about how much and what type of education to pursue, resulting in a labour market that operates cyclically with continual shortages and surpluses of skilled labour (Freeman, 1971).

Third, questions have been raised about how well prices can effectively regulate the market for education. In neo-classical economics, a perfectly competitive labour market regulated through prices provides the most optimal or efficient allocation of labour. However, some scholars have argued that allocation decisions also involve questions of equity – that is, who gains and who loses in any allocation decision – that cannot be addressed through efficiency criteria alone (Windham, 1988; Klees, 1989). Moreover, the price system is itself distorted because the costs of investing in education and the costs of using educated labour are subsidised by governments through such things as subsidies to students and educational institutions, wage restrictions (like minimum wages) and tax credits to firms. Finally, since education is said to provide a large number of private and social benefits, these benefits might not be reflected adequately through the existing price structure.

The impact of education on productivity and earnings

The final difficulty in assessing the pecuniary benefits of education is that those benefits may not be as directly attributable to education as human capital theory posits. In human capital theory, the economic benefits of education result from two effects: *i)* the effect of education on worker productivity; and *ii)* the effect of productivity on earnings. Both have been the subject of considerable research and debate.

Proponents of human capital theory have advanced several explanations as to how education enhances work productivity. They include arguments that education enables workers to work better with other (*e.g.* capital) inputs, to use information better on costs, and to deal better with "disequilibria" (Griliches, 1969; Schultz, 1975; Welch, 1970). Yet, except in the case of agriculture (Jamison and Lau, 1982), there is little direct empirical support for these propositions. Other empirical evidence directly challenges them by showing that education and other forms of human capital do not always improve productivity (Medoff and Abraham, 1981).

Competing models suggest that the relationship between education and productivity is more complex than posited by human capital theory. One alternative model is that of job competition, where individuals compete for available jobs that have different skill requirements, wages, and other characteristics (Thurow, 1975). Thus, workers' productivity and earnings result from the jobs that they hold and not directly from their education and training. As a result, there can be a "mismatch" between the skills and education level of individuals and those that are actually required to perform their jobs. This mismatch can have adverse effects on both productivity and earnings rather than the positive effects suggested by human capital theory (Rumberger, 1987; Tsang and Levin, 1985). Moreover, the mismatch can be qualitative as well as quantitative. That is, workers can have too little or, more typically, too much schooling for the jobs they hold, which is sometimes referred to as over-education or surplus schooling. Or they can have the wrong kind of schooling. For instance, Resnick (1987) argues that in the United States there currently exists a gap between the everyday, practical, real-world intelligence that is required in work and the formal, academic intelligence that is taught and measured in schools.

The primary difference between the human capital and the job competition models is one of emphasis. In the human capital perspective it is primarily the attributes of individuals, such as their knowledge and skills, that determine productivity in the workplace, whereas in the job competition perspective it is primarily the attributes of jobs, such as their skill requirements, that determine workplace productivity. A third perspective would be a truly interactive model, whereby attributes both of individuals and of their jobs determine workplace productivity and other outcomes (Rumberger, 1988). This latter perspective implies that three types of skills and abilities may influence workplace productivity:

a) the skills and abilities or human capital that workers bring into their jobs;
b) the skills and abilities that their jobs require as initially designed;
c) the skills and abilities that workers actually use to perform their jobs.

The latter may be influenced, in part, by workers using different strategies and skills to perform their jobs (Scribner, 1986) or by informal organisation of the workplace where workers share knowledge and skills (Darrah, 1990).

Another challenge to the human capital model concerns the relationship between productivity and earnings. Human capital theory, based on the tenets of neo-classical economics, assumes that workers' wages are directly proportional to the value of their marginal contribution to work output. Yet some empirical research has shown that wages are not always proportional to productivity (Gottschalk, 1978). Rather, wages can be determined by "sociological" factors, such as union bargaining, and not simply economic efficiency (Piore, 1973).

Implications

The foregoing discussion suggests that the relationship between education outcomes and economic outcomes is much more complicated and less straightforward than human capital theory suggests. This conclusion is supported by several other recent critiques of the economics literature (Blaug, 1985; Klees, 1989). Together, these critiques suggest two important limitations on the ability to assess the economic outcomes of education.

First, economic outcomes are derived in the labour market and thus are influenced, at least in part, by factors of both supply and demand. As such, a secular change in either supply or demand can affect the economic outcomes of education. For example, in the United States a large increase in the supply of college graduates into the labour market because of the baby boom depressed graduates' earnings in the 1970s (Freeman, 1979; Welch, 1979). The same baby boom was also responsible for increasing youth unemployment in the late 1970s and early 1980s (Freeman and Wise, 1982). More recently, the increased use of new technologies and the growth of the service sector are said to have increased the demand for skilled and educated labour (Rumberger and Levin, 1989).

How the supply and demand of educated labour interact to produce economic benefits is difficult to predict:

> "The evidence on the over-time behaviour of the returns to education is consistent with what Tinbergen (1975) described as the 'race between education and technology', namely, educational expansion shifts the supply curve to the right, whereas technological advances also shift the demand curve for skilled persons in the same

265

direction. Therefore, the eventual observed intersection between supply and demand over time is an empirical question." (Psacharopoulos, 1989, p. 227)

The second limitation in trying to assess the economic benefits of schooling is that the labour market itself may work imperfectly and distort the observed economic benefits of education. A number of challenges and criticisms have been raised about the conventional view of the economic outcomes of education based on human capital and neoclassical economic theories. These include questions about the content and determinants of education and skills, barriers to competition due to internal and segmented labour markets, the availability and distribution of information, and the impact of education on worker productivity and earnings. These challenges suggest that the economic outcomes most commonly attributed to education, such as employment and earnings, may provide quite imperfect and distorted information about such questions as how well the education system is meeting the needs of the economy, how much education is contributing to economic growth, and other questions that education indicators are frequently supposed to help address.

A conceptual model for developing labour market indicators

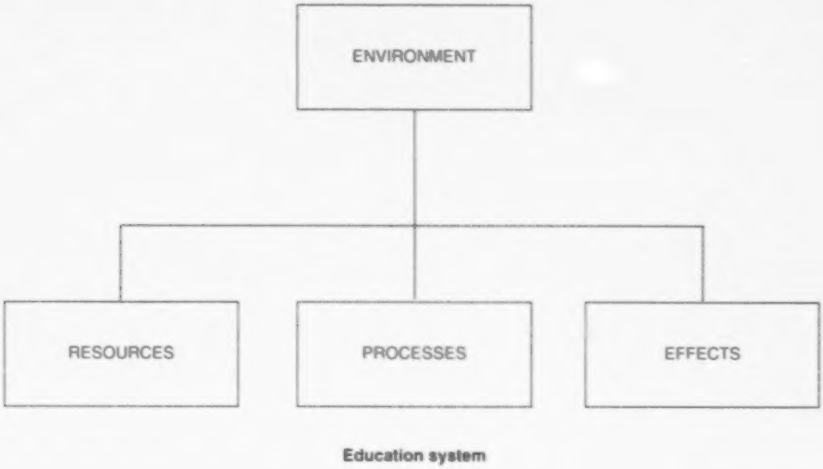
The previous discussion points out that while there is a substantial research literature to support the linkage between educational outputs and a variety of economic and social outcomes, there are also important limitations to inferring a direct, causal linkage. These limitations do not negate the value of trying to monitor and gauge the economic outcomes from schooling, but they suggest two cautions in attempting to do so. First, it is important to acknowledge from the onset the inherent limitations of any indicator system to monitor successfully the operation of a complex, social system. Second, to address at least partially this first concern, the indicator system should be based on a robust, conceptual model of how the economic system works and of how it is related to the outputs of the education system. This section, therefore, proposes a conceptual framework that could be used as a basis for developing labour market indicators of school performance.

The conceptual model that is proposed is not taken from any single source or analysis of the economic system. Rather, it attempts to pull together information from a variety of sources – both conventional views and more heterodox perspectives – to formulate a model of how educational outputs are translated into economic and social outcomes. Like any model, it should only be considered a simplification of a complex system. As such, it does not attempt to identify all the factors that can influence the broad array of economic and social outcomes associated with education, nor does it attempt to model accurately the causal relationships among these factors. The conceptual model expands a basic model of the schooling system and its environment, which is shown in Figure 13.1.

In Figure 13.1 the outputs of schooling flow back into the external environment, which includes the economic, social and political systems. In order better to identify labour market outcomes and their relationship with the school system, it is necessary to expand the linkage between educational outputs and the social environment.

This expanded conceptual model is illustrated in Figure 13.2. Its major components fall into four categories:

Figure 13.1 **General framework for the definition of international indicators of education systems**



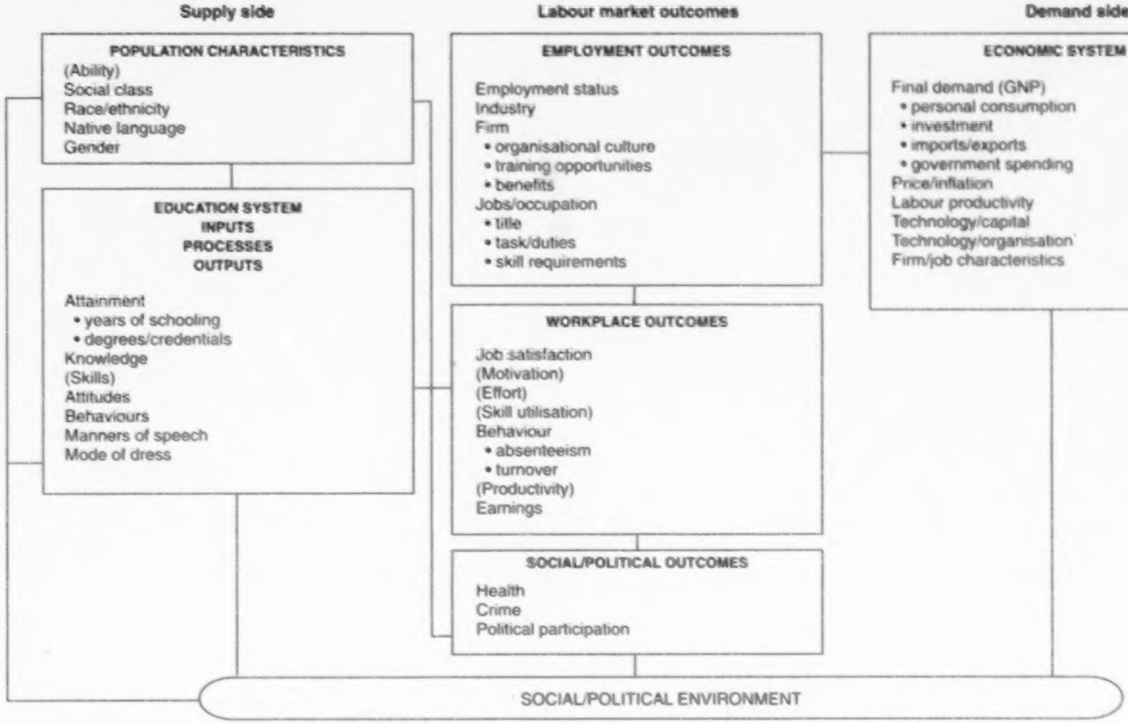
a) supply-side components, such as population characteristics and the education system;
b) demand-side components, which consist of the economic system;
c) labour market outcomes, which result from the interaction of the supply-side and the demand-side components;
d) the social/political environment, which initially affects both the supply-side and demand-side components and is then affected by the outcomes of the education system and the labour market.

The expanded conceptual model shows the relationship among the major components as well as a number of specific attributes and characteristics within each component. These specific attributes could be used as a basis for an indicator system to monitor the economic and social outcomes from schooling. In some cases these attributes are well known and easily identified with existing data and information. In other cases, they are perhaps well recognised, but not easily identified with existing data. The latter category of attributes or characteristics are shown in parentheses. Some of the key components of the model are described below in more detail.

Educational outputs

The model incorporates a broad array of educational outputs based on the proposition that a variety of these outputs are important in generating economic and social outcomes. These include not only achievement and skills, as identified in the human

Figure 13.2 **Conceptual model linking educational outputs to labour market outcomes**


Supply side
Labour market outcomes
Demand side
POPULATION CHARACTERISTICS
(Ability)
Social class
Race/ethnicity
Native language
Gender
EDUCATION SYSTEM
INPUTS
PROCESSES
OUTPUTS
Attainment
• years of schooling
• degrees/credentials
Knowledge
(Skills)
Attitudes
Behaviours
Manners of speech
Mode of dress
EMPLOYMENT OUTCOMES
Employment status
Industry
Firm
• organisational culture
• training opportunities
• benefits
Jobs/occupation
• title
• task/duties
• skill requirements
WORKPLACE OUTCOMES
Job satisfaction
(Motivation)
(Effort)
(Skill utilisation)
Behaviour
• absenteeism
• turnover
(Productivity)
Earnings
SOCIAL/POLITICAL OUTCOMES
Health
Crime
Political participation
ECONOMIC SYSTEM
Final demand (GNP)
• personal consumption
• investment
• imports/exports
• government spending
Price/inflation
Labour productivity
Technology/capital
Technology/organisation
Firm/job characteristics
SOCIAL/POLITICAL ENVIRONMENT

Note: Attributes enclosed in brackets are not readily observed or frequently measured.

capital literature, but also attitudes, values, behaviours, manners of speech, and even modes of dress that may be important in securing a job and succeeding in the labour market.

These outputs are influenced, in part, by an array of individual attributes, such as ability, social class, race and ethnicity, native language, and gender. The degree to which educational outputs are related to initial individual attributes is highly debatable, with some arguing that education provides a powerful mechanism for overcoming initial disparities among individuals and others arguing that the education system is a primary mechanism for perpetuating and legitimising these disparities (Carnoy and Levin, 1985).

Labour market outcomes

The labour market is the primary mechanism for allocating individuals with various characteristics to firms and jobs where they are required. The demand for labour is influenced by a host of factors, including the size and composition of final demand in the economy, prices, labour productivity and technology, which includes both capital and production processes or workplace organisation. The supply of labour is influenced by the size and characteristics of the population, which includes both initial population characteristics and those "produced" by the education system. The model does not resolve the issue of whether the education system actually "produces" the various characteristics associated with school, as in the human capital perspective, or simply "screens" individuals who already have those characteristics in the first place. But it does recognise that both educational outputs and individual attributes are important in determining the kind of job a person gets.

For instance, individuals of different social origins, race and ethnicity, native language, and gender may have quite different labour market experiences, even if they possess the same educational qualifications. This proposition is supported by a large body of research that shows widespread differences in the economic outcomes or benefits of various social groups (Weis, 1988). It also suggests that economic and social indicators need to be collected separately for various social groups in order to document these differences accurately (Oakes, 1989).

Employment outcomes

The primary result of the labour market process is to secure employment. A person's employment status indicates whether he or she secured a job. In addition, the type of employment secured can be further identified in terms of the sector where one is employed (public versus private), industry, firm, and specific job or occupation. Both the firm where one is employed and the job or occupation held can have several characteristics that are important in understanding the nature or quality of the employment secured. These qualitative aspects of employment may be very important to understand the qualitative "fit" between the needs of the organisation and the job, and the amount and types of education and training that workers bring into their jobs (Rumberger and Levin, 1989).

To illustrate, firms can differ widely in the type of organisational climate and structure in which they operate. Recent research literature suggests that new organisational forms are emerging in the United States, in part based on the Japanese model, that are less hierarchical and require more decision-making at lower levels of the organisation

(Hayes and Jaikumar, 1988). If that is the case, then workers may need different types of skills, knowledge, and even attitudes to work effectively in such organisations.

Workplace outcomes

A second set of economic outcomes results from initial employment outcomes. These are labelled workplace outcomes because they are derived from the job and work situation where one is employed. In the proposed scheme, these workplace outcomes are not only influenced by the setting in which one works, but also by the educational and individual characteristics that workers bring into their jobs. In other words, education influences workplace outcomes in two distinct ways – first, through its effect on the types of jobs that workers get and, second, through its direct impact on a number of interrelated workplace attitudes and behaviours (Rumberger, 1988).

This perspective is derived from a growing research literature that suggests worker productivity is influenced by both capacity to do work (human capital) and willingness or effort to perform it (Tsang and Levin, 1985). In this view, education can improve workplace productivity and earnings by increasing the productive capacity of workers, but it can detract from productivity if workers find themselves in jobs where their skills are not required or where the earnings, working conditions, or other job characteristics do not measure up to a worker's expectations (Rumberger, 1981). Recent research literature suggests that a "mismatch" between the characteristics of work and those of workers can have adverse effects on productivity and earnings (Rumberger, 1987; Tsang, 1987; Hartog and Oosterbeek, 1988).

Social and political outcomes

The final set of outcomes associated with education is related to labour market outcomes. These are social and political and include such variables as poverty, health, crime, and other outcomes discussed above. As in the case of workplace outcomes, they are influenced both by economic factors and by educational and individual attributes. Poverty, for instance, is directly related to the economic status that people attain in society – people who are unable to find work or only find low-paying jobs are more likely to be poor and require government support. Crime and health, at least mental health, may also be directly affected by a person's employment experiences. Other social outcomes, such as physical health and political participation, may be more directly influenced by educational and individual factors.

Development of economic indicators of education systems

Based on the foregoing conceptual model, it is possible to propose a set of indicators that could be used to measure the economic outcomes associated with education. This set attempts to measure aspects of the supply of educational outputs in the labour market, of the demand for those outputs, and a series of outcomes that result from the interaction of supply and demand. Each of these indicators would be relevant for understanding how well the education system is responding to the needs of the economy. They are described only at a conceptual level. Precise, operational measures would have to be developed based on other criteria, such as the availability of specific data (Shavelson *et al.*, 1987).

Seven indicators are proposed:

1. educational qualifications of the population;
2. employment status of the population;
3. characteristics of firms and jobs in the economy;
4. extent of "mismatch" between the qualifications of the population (No. 1) and the characteristics of firms and jobs (No. 3);
5. formal and informal training opportunities for workers;
6. workplace attitudes and behaviours;
7. earnings.

Educational qualifications of the population

This indicator would measure the supply of educated labour available in the market. It would measure the broad array of qualifications – the amount and type of credentials, knowledge, skills and other characteristics – that people bring into the labour market and that are important in generating economic, social and political outcomes. Of course, even measuring a broad array of qualifications by themselves would not indicate whether those qualifications were actually produced by the education system, as in the human capital perspective, or simply identified by the education system, as in the credentialling perspective. Although the overall educational qualifications of the population would be important to measure, those of recent graduates and school-leavers would be particularly valuable to ascertain trends in the production of such skills and qualifications.

Employment status of the population

This indicator would measure the proportion of the population that is employed at any point in time. It is a commonly used measure of economic and social prosperity since unemployment influences other economic and social outcomes, such as poverty and crime. Information on recent graduates and school-leavers would again be particularly useful since it would indicate whether these new labour market entrants were having problems in securing employment.

Characteristics of firms and jobs

This indicator would serve as a measure of the demand for workers with various types of qualifications. Although data on characteristics of firms and jobs are not commonly collected and difficult to use because they change over time, they have still been employed to get some indication of trends in the demands of work over time. Many governments, for example, use employment projections to forecast future needs in major occupational areas as a basis for promoting policies to address potential shortages. Although attempting to match education and training programmes precisely to the expected needs of the economy through the practice of educational planning has been widely criticised as unworkable (Klees, 1989), employment forecasts have still been used in a more general fashion to rally political support for reforming education (Johnston and Packer, 1987). In this way, more detailed information about the changing nature of work in firms and in jobs would be helpful to promote more specific reforms in the education and training system better to meet the needs of the economic system.

The mismatch between workers and jobs

An indicator or set of indicators should also be developed to measure any current mismatch between the number and qualifications of workers on the one hand (Indicator No. 1) and employment needs on the other (Indicator No. 2). This information would be particularly useful to ascertain for recent school graduates in order to see whether such leavers are fulfilling the current needs of the economy. In the United States, for example, there have been growing concerns that many students who leave before completing high school, and even graduates, are underskilled for the demands of work in the current economy (Rumberger and Levin, 1989). There are also concerns over shortages of specific types of workers, such as teachers and engineers, which such indicators could help identify.

Formal and informal training opportunities

The acquisition of economically useful knowledge and skills as well as the development of other useful attributes continues after formal school, both in formal training programmes provided by companies and through informal, on-the-job training. Indicators should be developed to measure the continued acquisition of such training to provide further information on the available stock of qualifications in the population.

Workplace attitudes and behaviours

In the conceptual model discussed above, a series of attitudes and behaviours were identified that have important implications for workplace productivity and earnings. They include job satisfaction, motivation, effort, and a number of behaviours such as absenteeism and turnover. Although many of these factors are not easily observed or measured, information on some of these attributes could be useful in attempting to identify problems associated with workplace productivity. In the United States, for example, there has been growing concern over a long-term decline in the productivity of American workers in the face of increasing qualifications of the workforce (Murnane, 1988). Information on more direct correlates of workplace productivity could help identify potential problems as to whether or how education is affecting workplace productivity.

Earnings

As discussed above, earnings are an important economic outcome in their own right and they are related to a number of important social outcomes as well. Moreover, they are easily measured and earnings data are widely available. Yet to operationalise this concept does present a number of problems, such as deciding exactly what type of earnings data to collect and how accurately to attribute observed differences in earnings to education versus other factors (Windham, 1988).

To move from a conceptual to an operational definition of economic indicators, a number of other factors must be taken into account. These will not be elaborated here since they have been identified and discussed in detail elsewhere. First, a number of other criteria must be applied in selecting specific, operational indicators (Shavelson *et al.*, 1987, p. 27; Murnane, 1987). Second, it is important that indicators are disaggregated for different populations and social groups in order to identify specific problems that such groups may be having in acquiring educational qualifications and achieving economic and social success (Oakes, 1989). A related consideration is that indicators be developed that can adequately ascertain issues of equity with respect to the distribution of economic outcomes in the various populations (Levin, 1976; Windham, 1988, pp. 624-627). Third, it is important to establish appropriate benchmarks and comparisons in order to interpret the indicators (Windham, 1988). Global economic benchmarks, such as aggregate unemployment rates for example, could be used to interpret whether unemployment rates for particular educational groups are relatively good or bad given the overall economic climate. In addition to global benchmarks, comparisons could be made over time, among groups or schools, or between economic benefits and the individual or social costs associated with generating those benefits (Windham, 1988).

Conclusion

Education indicators have been proposed as a means for monitoring the performance of the education system and as a mechanism for improving or reforming it. Since many recent reforms in the education system have been predicated on the need better to align the performance of the education system with the economy, indicators of economic outcomes have been proposed as a way to monitor that alignment better. This chapter described the rationale for focusing on economic outcomes as a measure of educational performance and some of the difficulties in trying to measure the linkages between educational outputs and economic and social outcomes. It also proposed a conceptual model to describe those linkages and some specific indicators to monitor various components of the model.

The overall conclusion of this analysis is that economic indicators can be useful in judging the performance of the education system, that nevertheless performs other functions besides the preparation of an adequately trained workforce. There is a substantial body of empirical and theoretical research that demonstrates a powerful linkage between educational outputs and a variety of economic and social outcomes. However, the linkage is less straightforward and more complex than is commonly perceived. Thus, it is very difficult to determine accurately the extent to which educational outputs are contributing to economic and social outcomes as opposed to other factors. In particular, since economic outcomes result, at least in part, from the operation of the labour market and the interaction of supply and demand, changes in the supply characteristics of the population in the form of their educational qualifications can have no predetermined impact on economic and social outcomes. Despite these limitations, economic indicators can still be used to monitor some of the key components of the education and economic systems and help determine the extent to which the two systems are aligned. More careful identification and measurement of these indicators can serve a useful purpose for measuring educational performance and promoting policies to align the two systems better.

Bibliography

BACHARACH, S.B. (ed.) (1990), *Education Reform*, Allyn and Bacon, Boston, MA.

BECKER, G.S. (1964), *Human Capital*, National Bureau of Economic Research, New York.

BEHRMAN, J. and BIRDSALL, N. (1983), "The Quality of Schooling: Quantity Alone is Misleading" *American Economic Review*, Vol. 73, pp. 928-946.

BLAUG, M. (1978), *Economics of Education: A Selected Annotated Bibliography* (3rd edition), Pergamon Press, Oxford.

BLAUG, M. (1985), "Where Are We Now in the Economics of Education?", *Economics of Education Review*, Vol. 4 (1), pp. 17-28.

BOWLES, S. and GINTIS, H. (1976), *Schooling in Capitalist America*, Basic Books, New York.

CALIFORNIA STATE DEPARTMENT OF EDUCATION (1989), "Performance Indicators in Response to Proposition 98" (mimeo), State Department of Education, Sacramento, CA.

CARNOY, M. and LEVIN, H.M. (1985), *Schooling and Work in the Democratic State*, Stanford University Press, Stanford, CA.

CATTERALL, J.S. (1987), "On the Social Costs of Dropping Out of School", *High School Journal*, November, Vol. 71, pp. 19-30.

COHN, E. and GESKE, T.G. (1990), *The Economics of Education* (3rd edition), Pergamon Press, Oxford.

DARRAH, C.N. (1990), "Skills in Context: An Exploration in Industrial Ethnography", Ph.D. Dissertation, Stanford University, Stanford, CA.

DENISON, E.F. (1979), *Accounting for Slower Economic Growth*, The Brookings Institution, Washington, D.C.

DOERINGER, P.B. and PIORE, M.J. (1971), *Internal Labor Markets and Manpower Analysis*, Lexington Books, Lexington, MA.

EUROPEAN ROUND TABLE OF INDUSTRIALISTS (1989), *Education and European Competence*, European Round Table of Industrialists, Brussels.

FREEMAN, R.B. (1971), *The Market for College-Trained Manpower*, Harvard University Press, Cambridge, MA.

FREEMAN, R.B. (1976), *The Overeducated American*, Academic Press, New York.

FREEMAN, R.B. (1979), "The Effect of Demographic Factors on Age-earnings Profiles", *Journal of Human Resources*, Vol. 14, pp. 289-318.

FREEMAN, R.B. and WISE, D.A. (eds.) (1982), *The Youth Labor Market Problem: Its Nature, Causes and Consequences*, Chicago University Press, Chicago, IL.

FUCHS, V. (1974), *Who Shall Live?*, Basic Books, New York.

GARDNER, H. (1983), *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books, New York.

GORDON, D.M., EDWARDS, R. and REICH, M. (1982), *Segmented Work, Divided Workers*, Cambridge University Press, New York.

GOTTSCHALK, P.T. (1978), "A Comparison of Marginal Productivity and Earnings by Occupation", *Industrial and Labour Relations Review*, Vol. 31, pp. 368-378.

GRILICHES, Z. (1969), "Capital-skill Complementarity", *Review of Economics and Statistics*, Vol. 51 (4), pp. 465-468.

GROSSMAN, M. (1975), "The Correlation between Health and Schooling", in Terleckyj, N.E. (ed.), *Household Production and Consumption*, National Bureau of Economic Research, New York.

HANUSHEK, E.A. (1986), "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature*, Vol. 24, pp. 1141-1177.

HARTOG, J. and OOSTERBEEK, H. (1988), "Education, Allocation and Earnings in the Netherlands: Overschooling?", *Economics of Education Review*, Vol. 7 (2), pp. 185-194.

HAVEMAN, R.H. and WOLFE, B.L. (1984), "Schooling and Economic Well-being: The Role of Nonmarket Effects", *Journal of Human Resources*, Vol. 19 (3), pp. 377-407.

HAYES, R.H. and JAIKUMAR, R. (1988), "Manufacturing's Crisis: New Technologies, Obsolete Organizations", *Harvard Business Review*, Vol. 66 (5), pp. 77-85.

HILL, C.R. and STAFFORD, F.P. (1974), "Allocation of Time to Preschool Children and Economic Opportunity", *Journal of Human Resources*, Vol. 9, pp. 323-346.

JAMISON, D. and LAU, L. (1982), *Farmer Education and Farm Efficiency*, Johns Hopkins Press, Baltimore, MD.

JENCKS, C. *et. al.* (1972), *Inequality: A Reassessment of the Effect of Family and Schooling in America*, Basic Books, New York.

JOHNSTON, W.B. and PACKER, A.E. (1987), *Workforce 2000: Work and Workers for the 21st Century*, Hudson Institute, Indianapolis, IN.

KLEES, S.J. (1989), "The Economics of Education: A More than Slightly Jaundiced View of Where we Are Now", in Caillods, F. (ed.), *The Prospects for Educational Planning*, International Institute for Educational Planning, UNESCO, Paris.

LEIBOWITZ, A. (1974), "Human Investments in Children", *Journal of Political Economy*, Vol. 82, pp. 111-131.

LEVIN, H.M. (1972), "The Costs to the Nation of Inadequate Education", Study prepared for the Select Committee on Equal Educational Opportunity, United States Senate, United States Government Printing Office, Washington, D.C.

LEVIN, H.M. (1976), "Educational Opportunity and Social Inequality in Western Europe", *Social Problems*, Vol. 24 (2), pp. 148-172.

LEVY, F. (1988), *Dollars and Dreams: The Changing American Income Distribution*, W.W. Norton, New York.

MATHIOS, A.D. (1989), "Education, Variation in Earnings and Nonmonetary Compensation", *Journal of Human Resources*, Vol. 24 (3), pp. 456-468.

McMAHON, W. and GESKE, T.G. (1982), *Financing Education*, University of Illinois Press, Urbana, IL.

MEDOFF, J.L. and ABRAHAM, K.G. (1981), "Are Those Paid More Really More Productive? The Case of Experience", *The Journal of Human Resources*, Vol. 16 (2), pp. 186-217.

MICHAEL, R.T. (1972), *The Effect of Education on Efficiency in Consumption*, Columbia University Press, New York.

MICHAEL, R.T. (1973), "Education and Derived Demand for Children", *Journal of Political Economy*, Vol. 81, pp. S128-S164.

MINCER, J. (1974), *Schooling, Experience and Earnings*, National Bureau of Economic Research, New York.

MINCER, J. (1989), “Human Capital and the Labor Market: A Review of Current Research”, *Educational Researcher*, Vol. 18 (4), pp. 27-34.

MURNANE, R.J. (1987), “Improving Education Indicators and Economic Indicators: The Same Problems?”, *Educational Evaluation and Policy Analysis*, Vol. 9 (2), pp. 101-116.

MURNANE, R.J. (1988), “Education and the Productivity of the Work Force: Looking Ahead”, in Litan, R.E., Lawrence, R.Z. and Schultze, C.L. (eds.), *American Living Standards: Threats and Challenges*, The Brookings Institution, Washington, D.C.

MURPHY, K. and WELCH, F. (1989), “Wage Premiums for College Graduates: Recent Growth and Possible Explanations”, *Educational Researcher*, Vol. 18 (4), pp. 17-26.

OAKES, J. (1989), “The Distribution of Excellence: Indicators of Equity in Precollege Mathematics, Science and Technology Education”, in Shavelson, R., McDonnell, L. and Oakes, J. (eds.), *Toward a System of Indicators of Precollegiate Science and Mathematics*, Rand Corporation, Santa Monica, CA.

OI, W.Y. (1962), “Labour as a Quasi-fixed Factor”, *Journal of Political Economy*, Vol. 70, pp. 538-555.

OECD (1989), *Education and the Economy in a Changing Society*, OECD, Paris.

OSTERMAN, P. (1980), *Getting Started*, MIT Press, Cambridge, MA.

PIORE, M.J. (1973), “Fragments of a ‘Sociological’ Theory of Wages”, *American Economic Review*, Vol. 63, pp. 377-384.

PSACHAROPOULOS, G. (ed.) (1987), *Economics of Education: Research and Studies*, Pergamon Press, Oxford.

PSACHAROPOULOS, G. (1989), “Time Trends in the Returns to Education: Cross-national Evidence”, *Economics of Education Review*, Vol. 8 (8), pp. 225-231.

PSACHAROPOULOS, G. and WOODHALL, M. (1985), *Education for Development: Analysis of Investment Choices*, Oxford University Press, New York.

RESNICK, L.B (1987), *Education and Learning to Think*, National Academy Press, Washington, D.C.

RIBICH, T.I. (1968), *Education and Poverty*, The Brookings Institution, Washington, D.C.

RUMBERGER, R.W. (1981), *Overeducation in the U.S. Labor Market*, Praeger Publishers, New York.

RUMBERGER, R.W. (1983), “The Influence of Family Background on Education, Earnings and Wealth”, *Social Forces*, Vol. 61, pp. 755-773.

RUMBERGER, R.W. (1987), “The Impact of Surplus Schooling on Productivity and Earnings”, *Journal of Human Resources*, Vol. 12 (1), pp. 24-50.

RUMBERGER, R.W. (1988), “Assessing Work Skills: Conceptual and Methodological Issues”, *Report 89-CERAS-14*, Stanford University, Center for Educational Research, Stanford, CA.

RUMBERGER, R.W. and LEVIN, H.M. (1989), “Schooling for the Modern Workplace”, in *Investing in People: A Strategy for Addressing America's Workforce Crisis: Background Papers*, Commission on Workforce Quality and Labor Market Efficiency, United States Department of Labor, September, Washington, D.C., Vol. 1, pp. 85-144.

SCHULTZ, T.W. (1961), “Investment in Human Capital”, *American Economic Review*, Vol. 51 (1), pp. 1-17.

SCHULTZ, T.W. (1975), “The Value of the Ability to Deal with Disequilibria”, *Journal of Economic Literature*, Vol. 13, pp. 827-846.

SCRIBNER, S. (1986), "Thinking in Action: Some Characteristics of Practical Thought", in Sternberg, R.J. and Wagner, R.K. (eds.), *Practical Intelligence*, Cambridge University Press, New York, pp. 13-30.

SHAVELSON, R., McDONNELL, L. and OAKES, J. (eds.) (1987), *Indicator Systems for Monitoring Mathematics and Science Education*, Rand Corporation, Santa Monica, CA.

SPENCE, M. (1973), "Job Market Signaling", *Quarterly Journal of Economics*, Vol. 87 (3), pp. 355-374.

TAUBMAN, P. and WALES, T. (1974), *Higher Education: An Investment and Screening Device*, McGraw-Hill, New York.

THUROW, L.C. (1975), *Generating Inequality*, Basic Books, New York.

TINBERGEN, J. (1975), *Income Distribution Analysis and Policies*, North-Holland, Amsterdam.

TSANG, M.C. (1987), "The Impact of Underutilization of Education on Productivity: A Case Study of the U.S. Bell Companies", *Economics of Education Review*, Vol. 6 (3), pp. 239-254.

TSANG, M.C. and LEVIN, H.M. (1985), "The Economics of Overeducation", *Economics of Education Review*, Vol. 4 (2), pp. 93-104.

U.S. DEPARTMENT OF LABOR, Commission on Workforce Quality and Labour Market Efficiency (1989), *Investing in People: A Strategy for Addressing America's Workforce Crisis*, United States Government Printing Office, Washington, D.C.

WAGNER, R.K. and STERNBERG, R.J. (1984), "Alternative Conceptions of Intelligence and their Implications for Education", *Review of Educational Research*, Vol. 54, pp. 179-223.

WEIS, L. (1988), *Class, Race, and Gender in American Education*, State University of New York Press, Albany, NY.

WELCH, F. (1970), "Education in Production", *Journal of Political Economy*, Vol. 78, pp. 35-59.

WELCH, F. (1979), "Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust", *Journal of Political Economy*, Vol. 87, pp. S65-S97.

WINDHAM, D.M. (1988), "Effectiveness Indicators in the Economic Analysis of Educational Activities", *International Journal of Educational Research*, Vol. 12, pp. 666-675.

BEST COPY AVAILABLE

Part Four

INTERPRETING INDICATORS FOR POLICY

Chapter 14

Indicators and Accountability at the School and Classroom Level

by

Michael Eraut
University of Sussex, United Kingdom

While the term "accountability" has been in common use in education for 15 to 20 years, most of the processes involved are much older. Indeed, many are taken for granted as features of the way schools operate. Accountability concerns the compilation, evaluation and discussion of results; some would also include any subsequent action. Thus, any process that contributes to the collection and judgement of or response to information can be said to have a role in accountability. However, this definition implies a degree of orderliness which is seldom, if ever, found in practice.

Transactions with an accountability dimension

Information may be collected by any party to an accountability relationship, not only by those formally called to account. Some may be collected deliberately, but other information is likely to be acquired haphazardly, with little attention to its reliability or validity. One has only to consider the variety of information a parent receives about a school to recognise that formally provided information constitutes a relatively small portion of the whole. The school itself may use informal occasions like open evenings and concerts to engender trust and bolster its image, while parents acquire much of their information from children's comments and the local gossip network. Curiously, it is often this least authenticated information which carries the greatest conviction. Before condemning this unsatisfactory state of affairs, however, it should be remembered that, although schools possess more authentic information than parents, they select, organise and present it through communications intended to convey a particular message or impression – as do other public institutions. Parental trust may depend on the image of the headteacher, personal contact with teachers, and opportunities for parents to see for themselves (Becher *et al.*, 1981).

The relationship between information and judgementis also very complex. Studies of professional judgement suggest that the selection and weighting of available information is strongly affected by the predispositions of the judge. This process may be conscious, when people search for evidence to support their existing views, or unconscious, when people filter what they read and see through cognitive and perceptual frameworks that affect the selection and interpretation of information. Sometimes this is reinforced by the way in which the information is communicated. Failure to include taken-for-granted information presents a distorted picture; and misunderstanding the interests and standpoint of the audience alienates and destroys credibility.

Even the response to judgement can be difficult to understand amidst the welter of communication events. Token response, where people accept a judgement or promise action and then apparently fail to make any changes, is a well documented phenomenon in the study of innovation. Less well studied are those situations where external views are resisted, yet nevertheless influence practice. Sometimes this results from deliberation after discussion. But on occasion, the effect is more subtle: putting a topic on the agenda increases its prominence in people's minds and hence causes them to give it more attention. This effect can result in changing the balance between different objectives competing for time in a curriculum or increasing the attention given by a teacher to the children of interceding parents.

The role of indicators in such transactions

Similar considerations apply to the use of indicators in accountability transactions. Indicators play an important role in people's thinking, but only a small proportion of what might be called "indicators in use" satisfy the technical criteria which social scientists regard as essential. Some indicators remain implicit, especially those affecting the interpretation of observation or informal conversation, while others appear to have no logical connection with what they are supposed to indicate. In studies conducted at the University of Sussex, United Kingdom, parents frequently used children's casual behaviour as an indicator of the priority given by their school to the teaching of basic skills. This appears to reflect a notion of the "good traditional school" as one which disciplines children both in and out of class; and it raises the complex psychological issue of whether the prominence in public discourse of indicators such as spelling tests is not largely symbolic. Good scores on spelling tests, which may have little to do with effective communication, symbolise the proper disciplining of the younger generation. At one level, they are not an indicator of effective teaching, so much as an indicator of effective socialisation into traditional values and also, perhaps, into a dependent, submissive role in society.

How, then, can the value of indicators for accountability transactions be judged? For a social scientist, the argument must surely hinge on whether an indicator improves the selection of evidence, using criteria such as fairness, validity and appropriateness, and whether an indicator improves the communication of evidence through being accorded the same meaning by all interested parties. For an educator, however, two further criteria need to be considered. First, there is the issue of whether the use of a particular indicator will increase or decrease the cost of the accountability process, especially the opportunity cost, since the allocation of people's time is usually a major problem. Second, there is the much more difficult question of whether, and under what circumstances, the use of an indicator will improve the quality of education through its effect on any subsequent action. This last issue is one of the major themes of this chapter.

Structures of responsibility

Cohen and Spillane (Chapter 16) examine variations in the structure of responsibility at national level and issues such as centralisation, professionalism and the tension between national debate and ideological conflict. All these factors reappear at local level but will not be discussed here. Rather, its aim is to explore how they affect accountability at school and classroom levels. District and local officials will be treated only as external agents to whom a school may have to account and from whom it may receive evaluative messages, instructions or resources.

This section is concerned with the extent to which schools, and individual teachers within schools, are responsible for policy as well as for practice. Which decisions are made externally and which are left to the school itself? To what extent are schools expected to render accounts of their stewardship, of what they have done under delegated authority and of how they have managed any further internal delegation? And to what extent are they simply accountable for implementing externally determined policies and for following external instructions? While there is considerable national variation, this chapter will assume that, *de facto* if not always *de jure*, schools have significant delegated authority. Hence, attention will be focused on accounts of schools rather than of national or regional policies.

The structure of responsibility within schools is also important for those who wish to understand the relationship between accountability procedures and educational change or lack of it. Most schools have three levels of decision-making: decisions made for the school as a whole, such as the timetable and a range of broad school policies; decisions made by teachers for their own classrooms, such as the grouping and seating of pupils; and decisions made at an intermediate level, such as a subject department, a year-group or a small cluster of subjects or year-groups. This intermediate level is strong in some countries and weak in others, where there is practically no "middle management". In general, policies about such matters as teacher and student behaviour are more likely to be formulated at school level, and policies about instruction at department or teacher level.

Several styles of decision-making can also be discerned, which again affect the structure of responsibility. These are the autocratic, managerial, collegial and permissive (or *laissez-faire*) styles. The autocratic style implies that most decisions are made at the top with relatively little consultation. The managerial style is increasingly characterised by a senior management team and delegation, with clearly specified briefs and job descriptions. The collegial style involves collaborative policy-making and a high degree of participation in decisions. The permissive style involves the greatest degree of decentralisation but without the internal accountability that goes with managerial delegation. Combinations of styles are common; an autocrat in some areas may be permissive in others; and a managerial style at school level may be accompanied by a mixture of autocratic, collegial and permissive styles at department or year-group level.

Instructional leadership is a recurring theme in research into effective schools, yet the concept of the principal as an instructional leader is relatively new in most countries. In some countries, significant responsibility for teachers' instructional performance lies with district inspectors, who still have the greatest influence on teachers' career prospects. In others, there is such a strong tradition of subject expertise that secondary school principals feel unqualified to judge on instructional matters and rely on their heads of subject departments. In many schools, there is a long tradition of principals, and even heads of department, confining their attention to administrative matters and keeping clear of the classroom. These matters are raised because the use of indicators is primarily associated with effecting instructional improvement; yet many schools lack the management capability to act on evaluative comment, even if they wanted to. The major policy dilemma is whether to develop a school's capacity to provide its own instructional leadership, or to attempt to improve instruction from outside the school. This decision has profound implications for accountability policy.

The location of responsibility for instructional leadership is intertwined with yet another accountability issue: that of how best to handle responsibility for student achievement. A moment's reflection suggests that the performance of an individual pupil may depend on many factors: resource factors (class size, availability of learning materials); curriculum factors (relevance, appropriateness, attainability); the pupil's prior knowledge and ability; the quality of teaching; and motivational factors arising from home, school and the pupil's own personal commitment to learning. Educators would argue that performance is best when there is partnership and a sense of shared responsibility for learning between the teacher, the parent, the child, the school and the wider education system. Yet many accountability procedures appear to destroy the development of such a partnership by attributing responsibility to one partner alone. This leads to the abdication of responsibility as the threatened teacher or school seeks to deflect the "blame" onto the "less able" child, the "unsupportive" home or the "system" which fails to provide the necessary resources. In these circumstances, constructive action is unlikely. It can even lead to teachers in "difficult" schools awarding themselves a "discount on performance" and setting standards that are too low. The conclusion may well be that evaluative messages are more likely to have positive effects if they are transmitted and reviewed in a manner that encourages partnership and acceptance of joint responsibility for educational outcomes. Any use of indicators will need to take this into account.

The teacher as a reflective professional

Cohen and Spillane (Chapter 16) draw attention to the distinction between education systems that rely heavily on inspectors as connoisseurs of the craft of teachers and those that attend to years of service and qualifications rather than actual performance. They rightly suggest that connoisseurs who rely on personal, tacit knowledge are likely to reject the use of indicators as failing to take sufficient account of the inherent complexity of professional judgement. The bureaucratic system may be more sympathetic to indicators, but is less likely to use them in a manner sufficiently sensitive to context and opportunity to achieve the desired positive effects. A third possibility, not mentioned by Cohen and Spillane, is that judgement of teachers' performance is the responsibility of the school itself. This will be taken up in the next section. Meanwhile, Cohen and Spillane's contrast between tacit, craft knowledge and a social science approach, potentially more oriented towards the use of indicators, is very important.

In an attempt to characterise the knowledge base for good pedagogy, Eraut (1989) drew attention to four types of process:

i) processes for *acquiring information*, ranging from deliberate searching to the almost intuitive reading of an emergent situation;

ii) *deliberative processes*, such as planning, decision-making and problem-solving;

iii) *routinised action* and *skilled behaviour*, which characterises much classroom teaching – intuitive yet following discernible patterns and still under some overall cognitive control;

iv) *evaluating and controlling* processes that concern, first, how professionals assess the impact of their actions and evaluate their personal practice and that of their organisation; second, how they make use of this information to modify or rethink their decisions, work-patterns and policies. Thus, the term "controlling" is used partly in the cybernetic sense of obtaining and responding to feedback. At an informal level, it involves daily decisions about what to do, reflecting on what has happened, and learning from experience.

The *teacher as craftsman* is principally occupied with process *iii)*, with a small element of process *i)* in its intuitive mode. There is considerable evidence to show that routinised action is central to coping in the classroom; but without a significant element of process *iv)*, routines gradually decay and cease to serve their purpose, while at the same time becoming difficult to change.

So, should the aim be to develop *teachers as social scientists* by emphasizing processes *ii)* and *iv)*? This is unrealistic, since it not only neglects the need for process *iii)* knowledge, but also fails to recognise that teachers work in environments where decisions have to be made instantaneously on the basis of insufficient information. The scope for prior deliberation is limited. Moreover, the accumulated body of research has made it abundantly clear that social science knowledge is not capable, or likely to be capable, of providing blueprints for action. Most of a teacher's practical knowledge is gained from experience, and it is in the interpretation of that experience that theory plays an important role.

This reinterpretation of the role of social science, combined with recent research into teachers' thinking, has led to the development, since the early 1980s, of the concept of the *teacher as a reflective professional*. This is based on the following set of assumptions:

- A teacher needs to have a repertoire of methods for teaching and promoting learning.
- Both selection from and adaption of methods in the repertoire are necessary to best provide for particular students in particular circumstances.
- Both the repertoire and decision-making process within it are learned through experience.
- Teachers continue to learn by reflecting on their experience and assessing the effect of their behaviour and their decisions.
- Both intuitive information-gathering and routinised action can be brought under critical control through this reflective process and modified accordingly.
- Planning and pre-instructional decision-making is largely deliberative in nature, as there is too little certainty for it to be a wholly logical process.
- These processes are improved when small groups of teachers observe and discuss one another's work.

Significant features of this concept are: *a)* its dialectic view of the relationship between theory and practice, which challenges the idea of prescription by external experts; *b)* incorporation of a theory of how teachers can change their practice; *c)* emphasis on process *iv)*, which could involve the use of indicators; *d)* its implications for teacher accountability.

Studies conducted at the University of Sussex have shown that teachers recognise three forms of accountability – moral, professional and contractual (Becher *et al.*, 1981). Moral accountability towards those affected by one's actions is a principle recognised by all humanity, and it involves teachers' relationships with their students, with parents, and with colleagues. Professional accountability is concerned with upholding the standards and ethics of a profession; and teachers describe themselves as accountable for this to colleagues, to their subject community and even to teachers in general. One important aspect of this, which overlaps with moral accountability, is self-accountability or accountability to professional conscience. Contractual accountability comprises accountability to the employer – the school itself, the school district or regional/state/national government – and those whom the employer appoints as managers, inspectors, or superintendents. The reflective professional concept enables those obligations to be formulated in greater detail so that their implications for professional conduct are clear. Thus, being a reflective professional implies:

- a moral commitment to serve the interests of students by reflecting on their well-being and progress and deciding how these can be fostered or promoted;
- an obligation to review periodically the nature and effectiveness of practice in order to improve the quality of management, pedagogy and decision-making;
- an obligation to continue to develop practical knowledge, both by personal reflection and through interaction with others (Eraut, 1991).
- an obligation to collaborate with other teachers, both in reflecting on one another's practice and in activities which contribute to the professional management of the school.

The professionalisation of school management

Since the 1970s attention has been increasingly paid to the school itself as a major focus for attempts to improve education. On the one hand, there is a growing research literature on effective schools which focuses on the characteristics of schools considered more effective than others according to a range of criteria (Mortimore *et al.*, 1988; Purkey and Smith, 1983; Renihan and Renihan, 1984; Reynolds and Cuttance, 1989); on the other, there is the school improvement movement which focuses on processes by which schools may improve their quality. Both have been given considerable support by the OECD, whose publication, *Schools and Quality*, stated:

> "No amount of external decision-making, control, and planning, let it be as enlightened and sophisticated as the best education authority can contrive, offers a guarantee that all schools will perform equally well. Each school has a life of its own that precludes conformity to anything like an exact norm of behaviour and success. By common consent, it is recognised that some schools perform better than others." (OECD, 1989, p. 125)

It then listed ten characteristics of effective schools, including: *a)* collaborative planning, shared decision-making, and collegial work in a frame of experimentation and evaluation; *b)* positive leadership in initiating and maintaining improvement; *c)* a strategy for continuing staff development related to each school's pedagogical and organisational needs.

The quality of school management is clearly a vital factor both in school self-improvement and in the effective implementation of externally sponsored innovation

(Fullan, 1982). Hence, it is no coincidence that the last decade has seen the development of increasingly thorough and sophisticated approaches to the training of principals and other senior staff. Recently, there has also been greater emphasis on instructional leadership than on administration. This signals the professionalisation of school management; and an important aspect of this process has been a growing acceptance of the notion of a school as a professional institution in which accountability, both internal and external, plays a central role.

The evolving model of a professional school is characterised by a series of cyclical development and evaluative processes, in which needs assessment (an evaluative process) gives direction to development which is later modified by formative evaluation. At school level there are the processes of formulating and implementing a school development plan and updating it, monitoring all areas of school activity, and occasionally conducting searching reviews. At teacher level there are both individual and collegial aspects of staff development, teacher appraisal, and the setting of personal targets. At student level the most flexible modes of teaching and learning involve agreement on objectives, selection of learning activities, providing feedback and recording achievement.

Many schools have begun to implement parts of this model, but few have succeeded in putting all of it in place. Why then give it so much attention? Because, at least in theory, it promises to solve some fairly intractable problems. First, by interlinking evaluative and developmental processes, it can ensure that development is properly based on need rather than on current fashion and that evaluation leads to appropriate subsequent action. Second, by careful involvement of external stakeholders in needs assessment and review, external accountability can be achieved without depriving a school of its efficacy, its dynamic, and its ownership of its own development. These are salient characteristics of successful schools which motivate teachers and students alike and enhance their capacity to respond to new challenges.

The alternative to developing the professional school is to develop new instructional systems in a few trial or experimental schools and then disseminate them to the others. This was the approach to innovation of the 1960s and 1970s. However, research on the implementation of such externally developed systems has indicated that an adaptive process is needed to fit any new system to the particular circumstances of each individual school and classroom; it has also indicated that teachers need not only training in the use of the new system but follow-up support in their own classrooms. Providing such support from outside the school is very expensive, and there is no guarantee of success. The most economical, and possibly most effective way of providing classroom support is to train teachers within the school to take on that advisory role, and to provide further external support for teacher-consultants. It is also worth noting that by the time the adaptive implementation of innovations by individual teachers has been catered for, a move has imperceptibly taken place, back to the reflective practitioner model. The more reflective practitioners are better implementers of external innovations, while those who do not reflect are unlikely to make optimal use of external support.

Where does this argument about professionalisation lead? First, the two central concepts – the *reflective practitioner* and the *professional school* – are still being developed; the aim is to construct effective accountability models for the future, not to patch up inadequate models from the past. Second, the relationship between external accountability and internal accountability is crucial. It is hypothesised that, except in extreme cases where strong external intervention is clearly necessary, the translation of evaluation into action is most likely to occur when internal accountability functions properly and when it incorporates the "external dimension" in a manner approved by the relevant stakeholders. This has important consequences for the use of indicators, because it implies that they must satisfy both internal and external criteria for good communication to take place; and this depends not only on technical criteria but also on how they are used, and in what internal and external contexts.

The role of indicators in classroom decision-making

Three important factors are frequently forgotten when the role of indicators in classroom decision-making is discussed. First, unlike most other accountability arenas, the classroom is characterised by a surfeit of information. Teachers receive more information than they can possibly process, and much of it inevitably passes unnoticed. In order to cope with this information overload, they have to develop selective attention. This results in a system of "indicators in use" whose operation is little understood. While some of these indicators are used explicitly and deliberately, others remain implicit and are used intuitively, with little conscious awareness of the criteria involved in their application. The problem is not whether to use indicators but which indicators to use, and how to bring their use under some kind of critical control.

Second, a great deal of assessment takes place in all classrooms, much of it informal and communicated orally. Such assessment, however, is not primarily intended to provide information to guide decision-making. It forms an integral part of the teaching process, providing both motivation and a final consolidation stage to a learning sequence. Psychologically, it closes rather than opens the study of a topic; and the recording and structuring of information about pupils' performance seems more "for the record" than for guiding forthcoming decisions. It is rare for assessment to be "tightly coupled" to classroom decision-making in the manner assumed by the proponents of diagnostic testing.

Third, the teacher is not the only decision-maker in the classroom. Students also are constantly making decisions. They may not always have the opportunity to choose their assigned activity, but they make short-term decisions about attention and medium-term decisions about the level and direction of their effort, not to mention decisions about how to tackle their work, all of which affect the learning process itself. Moreover, if the classroom arena is extended to include homework, even parents may be counted as proxy decision-makers on classroom matters.

Teachers make classroom decisions at three levels: that of the class as a whole, that of individual students and that of temporary or semi-permanent sub-groups of children within a class. The balance between these levels will vary according to the pattern of classroom organisation (*i.e.* the relative proportion of whole-class teaching, group work and individual work), the subject (reading, mathematics and art tend to be treated more individually than history or music), and the number of classes taught by any one teacher. Thus a specialist teacher who sees two hundred different students a week is more likely to think in terms of classes as whole entities than a generalist teacher who teaches only one class. For brevity, the focus will be only on the whole class and individual pupil levels.

Decisions at classroom level

Decisions about a whole class tend to concern pacing, approach and repetition or revision. What kinds of information influence these decisions? First, there is ongoing evidence of how the class behaves, both in general and in response to the teacher. The indicators in use will be those for attention and involvement, particular pupils often being treated as "markers". Second, there is evidence of understanding the nature of the tasks being set and the central concepts and ideas of the topic being taught. Sometimes this is revealed by vacant expressions or lack of participation, and sometimes it is ascertained by questioning. Evidence from written work tends to come later and is less likely to affect decision-making, unless there are clear signs of widespread failure to grasp some particular point. Otherwise the plan for covering the syllabus predominates. Short-term flexibility to sustain attention and respond to obvious problems of comprehension is regarded as good teaching; but if delays accumulate and seriously threaten completion of the syllabus, a mini-crisis develops. Teachers respond to this in different ways, but for most their accountability to cover the syllabus is the strongest pressure, possibly because this is their responsibility alone, whereas lack of learning is equally the "fault" of the pupils.

So, where is the scope for learning from experience? Where the teacher has a flexible repertoire, the success of activities and sequences will be noted, and less successful ones modified or replaced subsequently. Success is most commonly judged in terms of student responsiveness, but evidence of lack of understanding may also be noted for future consideration. The most likely response, however, to an indication that a particular topic was less well understood is to allocate it more time, hoping that performance will not suffer on those topics that have to be cut back in compensation.

In conclusion, it may be noted that behavioural indicators predominate over intellectual indicators, and process indicators over outcome indicators. The process indicators will be used intuitively and are unlikely to be subjected to any critical examination, unless there is some deliberate review involving either action research or observations by another teacher. The outcome indicators are more easily refined, but is there any need to do so? The problem lies not in the evidence – this is usually provided in abundance by student classwork, teacher-set tests and school examinations – but in the use to which it is put. The reflective study of student achievement, conducting item analyses or carefully scanning for areas of weakness, is not a regular habit of teachers. Thus, it is not the lack of indicators but the lack of regular review procedures in schools which limits progress in this direction.

Decisions at individual level

At individual level, the teacher faces two types of problem: how to collect and process information about individual pupils, and how to organise the class in a way that opens up possibilities for using that information. The latter problem is undoubtedly the greatest; there are enormous constraints on the extent to which individualisation is possible in crowded classrooms. But the former is more relevant to our topic. The teacher wants to know what each student knows and has recently learned, and to predict the likely outcomes for that person of various possible learning activities. Once more, research has shown that process evidence plays a greater part than might have been anticipated:

> ``The class teacher is not only able to inspect a child's work but also knows the setting in which it was done, its relation to previous and subsequent work, how much time it took to complete and how much help was given; and he or she has the added opportunity of discussing the work with the child. In going through children's exercise books with their teachers it is noticeable how much this contextual knowledge contributes to the judgements being made and how little can be deduced from the written page alone.'' (Becher *et al.*, 1981)

Such knowledge extends beyond written work to cover all other aspects of classroom interaction. Thus, a teacher's experience of a pupil is somewhat analogous to a series of film-clips in which many incidents are recorded, each of which is full of meaning. However, some of these incidents are likely to assume greater significance than others, according to their salience and their compatibility with previously held beliefs. Over time, the teacher constructs a view of the pupils, sometimes rather firm, sometimes provisional, which attempts to account for their behaviour.

This view will then influence, not always at a conscious level, the way the teacher interacts with that pupil in and out of class. Hence, the teacher's decisions are based not only on evidence of the pupil's work, but also on knowledge of the context in which it was done. However, the interpretation of this evidence is more complex still, because the teacher's thinking is also affected by his or her current view of the pupil, which has been developed through interactions and incidents involving that pupil over a considerable period of time.

The other aspect of teacher decision-making to be taken into account is the range of options available to any particular teacher. This depends not only, as suggested above, on the teacher's ability to organise a class so as to give more individual attention and to differentiate between students' needs, but also on his or her repertoire of possible learning activities and ways of responding to pupils' needs. Where this repertoire is limited, instructional decisions are confined to pacing and repetition. The decision-making process becomes sufficiently sophisticated to support using indicators only when the repertoire is large enough.

Four approaches to decision-making

In practice, four approaches to decision-making can be observed in classrooms:

- The automatic model is used whenever there is a fixed sequence of tasks. For example a pupil on a graded reading scheme finishes Book 4 and follows straight on to Book 5; or a pupil finishing an exercise in mathematics is told to go on to the next page of the textbook and do some more problems.
- The intuitive model is used all the time in classroom interaction, when a teacher responds to a pupil in a manner that is strongly influenced by how he or she regards that pupil but with little time to think about that pupil's need at that particular moment.
- The diagnostic model assumes that once appropriate evidence about a pupil's learning has been collected, the question of what to do next is resolved. It is an approach much advocated in debates about indicators, but which assumes that there is ``tight coupling'' between the diagnosis and the ``treatment''. The problem is not only that of the repertoire but also the uncertain state of pedagogic knowledge, which cannot provide that kind of ``scientific'' certainty.

- The deliberative model assumes that diagnostic knowledge and ideas for decision options have to be analysed and mulled over until an apparent "best fit" decision emerges. The approach is that of the reflective professional discussed previously.

The first two models cannot accommodate the use of indicators. The diagnostic model makes considerable use of indicators but tends to be rejected by teachers for failing to relate to their classroom reality. Not only is there suspicion of its claims of diagnostic value but it also prescribes spending an unacceptably high proportion of classroom time on testing. As a result, teachers tend to underestimate the contribution that indicators might make to the deliberative model, where precision is less important. Two areas are particularly worthy of attention: teachers' thinking about progression and teachers' constructs of pupils. Both are critical to decisions about what learning task to assign to a pupil and how best to provide a pupil with feedback.

The use of indicators presupposes an analytic approach to classroom decision-making and requires teachers to have a clear concept of what constitutes progress in the subject area concerned. This requires professional knowledge that is often quite scarce. While models of progression are common, though not uncontentious, in mathematics, they are scarce in history. History tends to be regarded in terms of covering a syllabus, with different degrees of student insight being ascribed to differences in aptitude rather than stages in the development of historical thinking. Where the teacher does have a model of progression, however, then the process of deciding where students have reached can be usefully assisted by indicators developed for that purpose. Such indicators need not be quantitative, such as scores on a progress test. They may take the form of checklists or special exercises such as concept mapping to reveal pupils' thinking about a scientific issue or topic.

Problems in decision-making

Three warnings should be heeded when indicators are used to assist decisions about the progress of individual students. First, if a deliberative model is followed, attention must also be given to decision options and the use of additional, less precise, sources of evidence. Second, the model of progression needs to be properly understood by the teacher. Third, the model must be built into the process of curriculum construction; if it is added afterwards, it will be at odds with the model assumed by the curriculum design.

For teachers' constructs of pupils, the use of indicators is a delicate matter when teachers judge the "ability" or "potential" of a student. The psychometric certainties of the immediate post-war years have crumbled away, leaving teachers vulnerable in this area and prone to being more tentative and less explicit in their public judgements. In some cases, this has driven teachers' views about pupils beneath the surface. Such views still influence classroom decisions about grouping, assigning work and giving feedback, but become less subject to reflective critical control when they cease to be explicitly discussed.

How then do teachers make judgements about their students' potential? They often assume that if the students are working hard, they are working at their potential. This unfortunately leads to a climate of underexpectation. Young children, in particular, will work hard at tasks that offer little mental stimulation; and research shows how adept students are at producing the minimum amount of work that satisfies the teacher, while managing always to look busy. They may also regard students' best work as an indicator of their potential, a practice not entirely immune to underexpectation. Or they may rely on process criteria, a more intuitive approach. Does the pupil ask intelligent questions, quickly understand an explanation, choose more rather than less challenging tasks? There is a danger that more extrovert behaviour will be more highly regarded. More dangerous still are tendencies to prejudge pupils' potential on the basis of deeply held assumptions about gender, class and ethnicity.

There is obviously some scope for the use of indicators that encourage and contribute to critical reflection about pupils' ability and potential, though again three warnings need to be heeded. First, indicators should not be given undue weight. Second, it is equally important to develop teachers' capacities to collect a wide range of evidence about students. Third, it is always useful to get a "second opinion" from another teacher, backed up if possible by other observations of pupils.

Finally, the issue of "pupils as decision-makers" presents similar problems. First, students need criterion-referenced information in order to make the maximum contribution to their own learning, but they rarely get adequate feedback. They tend to receive more information about what they have not learned than what they have learned, and even that often fails to suggest what they need to do in order to improve their performance. Second, pupils also construct views of themselves as learners which suffer from stereotyping and underexpectation. Even criterion-referenced information is used normatively as they compare themselves with fellow pupils. When personal messages are relatively rare, almost any piece of information is treated as an indicator. This leads to misunderstanding. What can be done? The communication of evaluative messages to pupils needs to be monitored and, if necessary, modified so that they have a positive effect on learning. Classroom discussion of learning and achievement has to move away from regarding learning as easy and failure to learn as reprehensible, which only creates feelings of inadequacy, towards regarding learning as a challenge and success as something to celebrate. The formation of pupils' attitudes towards different subjects and different aspects of their work needs to be monitored, for these also are affected by the nature of the feedback they receive. In general, whenever indicators are used, the way in which they are communicated is of vital importance. Moreover, it should be emphasized that any piece of information is likely to be used as an indicator regardless of its intended significance.

The role of indicators in internal accountability

Eraut (1978) distinguished three types of evaluative process in schools: general monitoring with follow-up inquiries and trouble-shooting; regular reviews incorporating needs assessment; and special investigations. One purpose of general monitoring is to check that nothing is seriously amiss; most of the responsibility is often delegated within the school to middle management and classroom teachers. The greatest attention is given to student behaviour, and reliance is placed on incidental information and occasional observation. Another purpose, for which procedures might need to be strengthened, is the monitoring of policy and performance standards, which could make some use of indicators. However, it should never be forgotten that monitoring can only indicate where there *might* be a problem needing attention, not where there definitely *is* a problem. More detailed evidence should always be sought without prejudice, as there are bound to be false alarms.

Regular reviews are an essential component of the professional school. Whereas monitoring is concerned with the implementation of existing policy and practice, reviews have to consider their overall impact and whether they are the best available. Thus, a review is both an evaluation of an existing situation and a consideration of possible alternatives. It is concerned not only with accounting for the past but even more with needs assessment for the future. Unlike monitoring, which is an ongoing process built into people's job descriptions, reviews are special activities involving the planned collection of new evidence and collaborative discussion of its implications. Any use of indicators will be part of the evaluation design, and findings will be considered together with other sources of evidence.

Special investigations are mounted when either monitoring or review has revealed the need for a major policy change. Even greater attention will be given to exploring policy alternatives, but in principle the use of indicators will be similar to that in reviews.

A system of regular reviews also has to collect data on a regular basis, for purposes of comparison. For comparisons over time, school-specific indicators that fit the school's policy priorities can be used. For comparison with other schools, common indicators are needed which may be less appropriate for any individual school. When accountability is internal, appropriate allowances can be made. However, there still remains the problem of getting sufficient contextual evidence about other schools or sample populations of schools to be able to compare like with like.

The evaluative processes of monitoring, review and investigation can be applied to three main objectives: the progress of individual pupils; various aspects of policy and practice; the work of individual teachers.

Most issues concerning individual pupils were discussed under classroom decision-making. But one is better debated here. Many schools currently review pupils' progress at the end of the year, after examinations, when they fill in official records. While such reviews contribute to revising or reaffirming the school's official definition of each pupil, they tend to have little effect on practice because the pupil is about to move on to another teacher. Mid-year reviews can contribute more to classroom decision-making and hence to the improvement of pupil performance, provided they are given that orientation. One possible model involves the review of individual pupils' progress by two teachers, followed by discussions with the pupils and their parents. The process is phased over two months in the middle of the year and results in action plans for pupil learning to which all parties contribute. The advantage of inter-teacher discussion is that it generates ideas and brings out unrecorded information about the pupils concerned.

Learning to undertake internal reviews of policy and practice is part of the process of professionalising school management. Much guidance can be obtained from the literature on programme evaluation, but it tends to become a more expensive and time-consuming operation than schools can afford and must be adapted. The whole process of managing such reviews within the school is also of crucial importance (Eraut, 1984). When properly managed, teacher ownership of the process is strong, morale is improved, and the chances of successfully implementing recommendations are high. When badly managed, teachers are alienated, conflict is generated, and the chances of effective subsequent actions are minimal. The negotiated use of performance indicators is a particularly sensitive issue, because it is where teachers usually feel most vulnerable. The more accountability is viewed as a professional approach to improving quality rather than an elaborate exercise in attributing blame, the better the effect.

The same consideration applies to evaluating the work of individual teachers. Teachers often feel trapped by the problems of attending to the individual needs of large numbers of pupils, whilst implementing school policy and attempting to adapt to new demands and new situations. Since extremely poor performance of the kind which might lead to dismissal is quickly detected by even the crudest monitoring system, there is little point for teacher evaluation procedures to have other than a strong developmental orientation. It was previously argued, with respect to the reflective professional, that teachers were accountable for their own continuing development as well as their current practice. The school's management has to promote and support this process. An appropriate appraisal system both gives positive feedback and provides an arena for negotiating a balance between the development needs of the individual teacher and those of the school. An appropriate staff development system needs to build in opportunities for mutual observation as a stimulus to reflection and its possible extension into action research. Both processes may make some use of agreed indicators, and it is here in particular that process indicators are best tried and tested. For example, issues such as pupil participation, the quality of interaction, and the balance of time between different activities lend themselves to the use of indicators developed by classroom researchers. Such work will also provide valuable guidance for reviews of policy and practice. The relationship between classroom processes and pupil performance should feature regularly, by discussions among teachers as a group of professionals considering ways to improve their practice.

The role of indicators in external accountability

External accountability takes place on two levels; accountability for school policy and practice to a wide range of stakeholders and accountability specifically to parents for the progress and well-being of their children. Communication between schools and external audiences suffers from all the problems raised in the introduction to this chapter concerning the provision and reception of information and the way in which judgements are made on the basis of very limited evidence. Outside the school, information is so scarce that almost any piece of information, however atypical or invalid, is likely to be treated as an indicator, especially if it confirms the previous opinions of the audience. Since much of this information lies outside the school's control, this presents a major public relations problem. Recent evidence of national and local debate suggests, first, that valid indicators do not drive out invalid indicators and, second, that positive messages cannot counter negative ones in any context where this does not suit the interests of the press or the politicians. This makes it difficult for serious proposals about external accountability based on proper use of valid evidence to carry much weight. However, this does not mean that the effort to develop procedures that may eventually acquire the necessary credibility should be abandoned.

What kind of action may be expected to lead to an improvement in quality at the school level? What can an external audience do about a school? First, and most important, it can engage in dialogue. Where both parties are willing to listen and respond, this is clearly the simplest and most appropriate course of action. Second, the community can stop using a school with a bad reputation. But there may not be an alternative school within reach, and transferring a child from a school is exceedingly disruptive. Third, lay people can complain to the appropriate authorities. What then can the authorities do?

They also are likely to engage in dialogue. The alternative options are all very drastic – changing the principal, sending in a professional team to sort the school out, closing the school – and unlikely to be followed unless there is a major crisis. Thus under normal circumstances, external accountability is essentially about two things: the collection of evidence and procedures for discussing that evidence.

The distinction between monitoring, reviews and investigations made in a previous section is still pertinent. Education authorities are likely to use some form of external monitoring. This will typically involve receiving community comments at random, occasional visits by officials and using a small number of performance indicators such as results in public examinations, occasional standardized tests, truancy rates and staff turnover. These indicators may be used to look at trends over time and to make comparisons with other schools. With external authorities, however, it is usually more difficult to contextualise this evidence and to convince them of the need to make further inquiries before jumping to premature conclusions. Where the school's internal monitoring is already providing the necessary information and reviewing it, there will be greater scope for constructive dialogue. Since schools cannot get comparative information without general agreement on certain indicators, an agreed policy between schools and authorities is needed. While authorities can, and often do, impose policies of this kind, the schools themselves are then less likely to make use of the information, thus defeating the main purpose of the policy.

Reviews and investigations can take three forms: internal, external, and internal with some external participation. Since schools can easily become isolated from a range of views and policy options, some external involvement is normally desirable. However, external reviews run the danger of contravening the principle of delegation. External investigations such as inspections are necessary when there is major concern about a school. For reasons of cost and comprehensiveness, internal reviews with external participation ought to play a major role in external accountability. There are several modes of participation, which are appropriate for different kinds of stakeholders. Some clearly require access to external professionals. These modes include:

- receiving a report and discussing its findings;
- providing evidence and opinion through interviews, statements or questionnaires;
- contributing questions to the agenda of issues under investigation;
- collecting evidence by observation or interview;
- suggesting additional policy options and providing information about other schools;
- assisting with data analysis and writing the report;
- auditing the report and providing independent professional comment.

This external participation serves two purposes. It improves the quality of reviews and it adds to their external credibility. It may increase the range of indicators used, but will not usually change their use.

External inspections of a routine kind are unlikely to make much use of performance indicators, but they may be concerned with process indicators linked to policy implementation. As Cohen and Spillane (Chapter 16) noted, inspection systems are largely based on connoisseurship. However, when inspectors are called in to make a special investigation in a crisis situation, premium placed on credibility may cause them to make greater use of indicators. Numbers are often easier to defend in public arenas than expert professional judgement.

Finally, there is external accountability to parents. At school level parents are one of the most important external audiences. They largely determine the school's reputation and initiate informal comments and formal complaints which contribute to external monitoring by education authorities. While it is not appropriate for parents to engage in formal monitoring, their participation in reviews is surely desirable. The first two modes listed above are particularly appropriate. Parents appreciate receiving almost any kind of information and their views on issues and priorities should be included in most reviews. Parents' concern, however, is with the progress of their own children. Their natural anxiety is often misinterpreted by teachers who see criticism in the simple urge to know what is going on. Hence, teachers tend to react defensively to proposals for parent involvement, as if their competence were somehow being threatened.

Although parents acquire some information about their children's progress by reading their books and talking to them, the formal mechanisms of written reports and teacher-parent interviews are of crucial importance. In general, it is assumed that parents have a right to know about their children's progress and that teachers have an obligation to tell. But how much? Reports and oral accounts differ greatly in detail. The use of indicators may seem an obvious way of improving such accounts and is usually welcomed by parents. But teachers are very divided on this issue. Quantitative indicators such as test-scores, class rank or marks on school examinations are most likely to be used when:

- decisions on differentiation are imminent, when they play an important role in explaining or justifying the school's decision or advice;
- teachers use marks and class rankings throughout the year as a means of motivating and controlling their class;
- the school or individual teachers believe parents have the right to be given "accurate" information about their children's relative achievement.

Indicators are most likely not to be used if information is not readily available because no tests or examinations have been administered or if teachers believe any of the following:

a) The indicators will not give a valid view of student achievement, because they are either too unreliable or too narrowly focused.

b) Indicators will be misinterpreted by parents who do not understand the margin of error or the likelihood of changes in relative performance over time.

c) Precise information might discourage both parent and child, lower confidence and have an adverse effect on future performance.

d) Parents might respond angrily to unwelcome information.

Argument *a)* depends on the choice of indicators and how information is presented to parents. A sufficient range is needed to demonstrate the diversity of children's achievements in different aspects of their work. Arguments *b)* and *c)* show the unacceptable side of teacher paternalism but still convey genuine concern. Two developments would help: devising modes of presentation which illustrate both the margin of error in any figure and the probability of shifts in children's relative performance over time; and developing a reporting system that stresses what children *have* achieved and strikes an appropriate balance between criterion-referenced and norm-referenced information.

Argument *d)* indicates lack of confidence in handling parents. The short-term worry about the threat to the powerful norm of "peaceful coexistence" dominates the long-term danger that parents will suddenly find out that their child has been achieving much less than they thought. Parents consistently confirmed that they desire to be told the truth, however bad; many teachers stressed that when achievement was poor they needed "to dress it up a bit". As a result, many accounts emerge so smothered in dressing that the central message is indiscernable. The notion that anger might be a reasonable emotional reaction and that they should ride out the storm, rather than seek to avoid it, does not occur to many teachers. In any case how could this possibly be done in a semi-public classroom in five to ten minutes?

Regrettably, teachers, and especially headteachers, feel uncomfortable when different teachers' accounts to parents are inconsistent, because it makes teachers vulnerable to criticism. This can be prevented either by using vague language – a renowned feature of written school reports – or by developing a school definition of the child. The school's long-term task is to get parents, in a succession of parent evenings, to accept the school's definition of their child, whilst still retaining their trust and confidence. An inherent attraction of firm quantitative data on student achievement is that, when added to an already developed reputation, it implicitly attributes the major responsibility for performance to the pupil alone.

Conclusion

The wheel has come full circle. The discussion has moved from the possible use of indicators to prevent teachers' expectations from becoming too dominant an influence on thinking about pupils' potential to the misuse of indicators to justify the teacher's view of the pupil and "calm" the parents. But is this dilemma not inherent to all information? The quality of information cannot protect it from misuse. Education in the proper transmission, reception and interpretation of information will always be needed.

Bibliography

BECHER, T., ERAUT, M. and KNIGHT, J. (1981), *Policies for Educational Accountability*, Heinemann, London.

ERAUT, M. (1978), "Accountability at School Level – Some Options and their Implications", in Becher, T. and MacLure, S. (eds.), *Accountability in Education*, NFER/Nelson, Windsor and Slough, pp. 152-199.

ERAUT, M. (1984), "Institution-based Curriculum Evaluation", in Skilbeck, M. (ed.), *Evaluating the Curriculum in the Eighties*, Hodder and Stoughton, London, pp. 54-63.

ERAUT, M. (1989), "Mid-career Professional Education", paper given to Greek Education Society, Athens, October.

ERAUT, M. (1991), "Teacher Accountability: Why it is Central to Teacher Professional Development", in Fessler, R., Kremer-Hayon, L. and Vonk, H.C. (eds.), *Teacher Professional Development*, Falmer, Basingstoke.

FULLAN, M. (1982), *The Meaning of Educational Change*, Teachers' College Press, New York.

MORTIMORE, P., SAMMONS, P., STOLL, L., LEWIS, D. and ECOB, R. (1988), *School Matters: The Junior Years*, Open Books, Salisbury.

OECD (1989), *Schools and Quality: An International Report*, OECD, Paris.

PURKEY, S. and SMITH, M. (1983), "Effective Schools: A Review", *Elementary School Journal*, Vol. 83, pp. 427-452.

RENIHAN, F.I. and RENIHAN, P.J. (1984), "Effective Schools, Effective Administration and Effective Leadership", *The Canadian Administrator*, Vol. 24 (3), pp. 1-6.

REYNOLDS, D. and CUTTANCE, P. (eds.) (1989), *New Directions in School Effectiveness and School Improvement*, Cassells, London.

Chapter 15

Interpretation and Analysis of Cost-per-Pupil Indicators in International Comparisons of Education Systems

by

Nico P. Mol
Faculty of Public Administration and Policy,
University of Twente, the Netherlands

and

Frans Kaiser
Centre for Higher Education Policy Studies,
University of Twente, the Netherlands

In comparisons of education systems, variances in indicator values often receive different interpretations. These diverging interpretations may have incompatible policy implications and thereby undermine the possibility of usefully applying the results of comparative research in order to improve these systems.

This predicament has been recognised from the start of the CERI project on International Education Indicators, and the necessity to prevent ambiguous meanings has been repeatedly stressed. Any attempt to compare education systems by means of indicators will require a conceptual framework. As Nuttall (1992) states:

> "To compensate for the unidimensional nature of each indicator, it is necessary to build a system of indicators, that is, a coherent set of indicators that together provide a valid representation of the condition of a particular education system, not just an *ad hoc* collection of readily available statistics." (p. 15)

Coherence among indicators

However, the ideal system of indicators will never be more than partially attainable. Two main obstacles stand in the way. First, as van Herpen (1992) shows in a review of the literature on models of education, systems of indicators may be specified in many different ways, since every conceptual model is linked to some underlying theory and numerous theories of education have been proposed. Precise interpretation of variances in indicator values thus depends on the theory chosen when the elements of the system are specified. Thus, attempts to derive one generally endorsed conceptual framework would

BEST COPY AVAILABLE

be vain, since different perspectives on education will call for different specifications of the model. Second, even if some theoretical specification were adopted, and equally relevant alternatives discarded, a fully valid and accurate set of indicators could not be produced, due to problems of measurement.

For both these reasons, the conceptual framework to be adopted in the INES project will not properly be a systems model. It will only be a loosely structured framework, building on the general systemic scheme depicted in Figure 15.1.

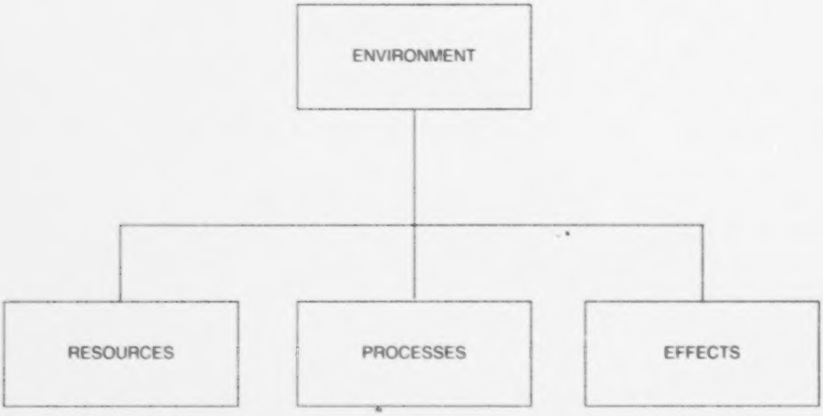
While such a framework will make possible an overview of the multitude of indicators proposed, thus allowing some analysis of the relationships among them, it will not avoid the predicament caused by diverging interpretations of observed variances in the indicators.

If ambiguities cannot be prevented because a single coherent model cannot be constructed which would allow a precise meaning for the indicators to be inferred deductively, frameworks must be developed to interpret indicators unambiguously in an "inductive" way. A starting point could be a specific indicator supposed to be relevant for explaining differences between education systems. Then, some composite set of variables conceptually related to this indicator may be elaborated to bring out the "hidden meaning" of its variances.

Heuristic approach to interpretation

This chapter explores two major issues involved in this heuristic (inductive) interpretation process. The first is the interpreting of indicators in a manner consistent with the underlying data. The second is the conditions which a set of variables has to meet in

Figure 15.1 **General framework for the definition of international indicators of education systems**



order to explain observed variances. These two issues cannot be treated independently. The interpretation of observations for an indicator will determine the selection of the variables needed to explain its variance. A precise and detailed analysis of observations is therefore necessary before any composite set of variables can be constructed. Previous research has often neglected this step. Researchers usually start by choosing a set of variables intended to explain observed variances in indicator values. The question of how the set is chosen is either not addressed or not treated systematically. Because different authors introduce different sets of variables, they may thereby propose diverging or even contradictory explanations for the observed variances.

The example of a specific indicator, cost or expenditure per pupil, is presented below. Many authors have discussed the meaning of this indicator and have investigated its relationships with other variables. More than any other indicator in economic research on education, the meaning of the cost per pupil variable depends on the underlying context.

This chapter first surveys the different approaches encountered in the research literature. Second, it presents a framework for systematic interpretation and analysis of the indicator. Finally, this framework is applied to the interpretations of the indicator in the research.

A review of the literature

It is far beyond the scope of this study to consider all the investigations dedicated to the cost per pupil indicator. Hence, this review is limited to a number of major contributions in the field.

Two obviously diverging approaches to interpreting the cost per pupil indicator can be distinguished. In the first, costs per pupil are supposed to reflect the "efficiency" of the services provided by an education system. From this perspective, the question asked (and strongly debated) is how this efficiency might be explained by characteristics of the system (*e.g.* size of classes and schools). In the second approach, costs per pupil are interpreted as a measure of the "effort" expended on education. From this perspective, the question (strongly debated as well) is the extent to which pupils' achievements can be attributed to educational expenditure instead of pupils' family background, for instance.

Economies of scale

The first approach is especially found in studies concerning economies of scale. In these studies, costs per pupil are related to institutional characteristics, *e.g.* the size of the school or school district. In the studies by Hirsch (1966), Riew (1966), Brinkman (1981) and McLaughlin *et al.* (1980), the size of the educational institution is not the only variable affecting costs per pupil; changes in enrolment (Hirsch, Riew), teachers' qualifications (Hirsch, Riew), population density (Hirsch), curricular diversity and complexity (Riew, McLaughlin, Brinkman), the wealth of the school district (Hirsch), the condition of the building (Riew), and teacher/pupil ratios (Bowen, 1981; McLaughlin) appear to be related to costs per pupil. However, little consistency can be found in the direction and extent of the relations between these variables and costs per pupil.

The confusion that might rise from this diversity in variables affecting costs per pupil is augmented by the finding that definitions of costs as well as of pupils differ across these studies. Hirsch, McLaughlin and Brinkman refer to costs as instructional costs, whereas Riew uses total expenditure of school districts. Pupils are measured either by average daily attendance rates (Hirsch, Riew) or as full-time equivalent enrolment rates (Bowen, McLaughlin).

Effort spent on education

The second approach manifests the same complexity. In most studies, expenditure per pupil appears to have no consistent correlation with pupils' achievement. In only one study (Sebold and Dato, 1981) is a positive correlation found. A wide and diverging set of variables other than costs per pupil is used to calculate the determinants of pupils' achievement. The main one is "family background", used by Coleman *et al.* (1966), Walberg and Fowler (1987), Hanushek (1989) and Sebold and Dato. Others are: influence of the peer group (Coleman *et al.*; Hanushek); size of the school district (Walberg and Fowler, Sebold and Dato); affluence of the school (Bowen); the structure of the compensation scheme for teachers (Stern, 1989); teacher/pupil ratios (Stern, Hanushek); innate learning capabilities, the influence of the curriculum and instructional methods (Hanushek); and entry level test scores (Sebold and Dato). The findings concerning the influence of these variables on student achievement are, except for family background, not very consistent.

Definitions of costs or expenditure as well as of pupils also differ across the studies. Coleman *et al.*, Stern, Hanushek, and Sebold and Dato confine themselves to educational expenditure, but Walberg and Fowler use a broader definition. The number of pupils is measured either in full-time equivalent enrolment rates (Coleman *et al.*, Hanushek), in headcount enrolment (Eicher, 1989), in average daily attendance rates (Walberg and Fowler, Sebold and Dato) or in daily student class hours (Stern).

It thus appears that there is little consistency in the use of the cost per pupil indicator and in the conceptual frameworks within which it is interpreted. There is thus a need for a systematic approach to the interpretation of this indicator.

Development of an interpretation scheme

On the one hand, variances in the costs per pupil among education systems seem to call for judgements about the relative efficiency of the systems. On the other, they have led to judgements about the relative levels of effort made for educating the population.

In each case, the cost per pupil indicator is placed in a different conceptual framework. Underlying the first approach is a systems model that sees the indicator as a "full cost" measure for a unit of production in education. Here, costs per pupil are seen in the light of an input-output ratio. The systems model underlying the second approach views the costs as an input or process indicator, to be related to outcomes of education as measured by pupils' achievement.

The fact that these frameworks are not always explicit represents a potential danger when using the indicator to compare education systems. In any systems model in which the indicator may be embedded, a particular set of *ceteris paribus* conditions must hold if valid conclusions are to be drawn from measured variances. For instance, the first approach requires assumptions with respect to participation and completion rates, which may or may not be relevant. The second approach may distort results by varying price levels of the inputs represented by the indicator, but the variations may be of no interest in an intended "full cost" confrontation.

In much previous research the conceptual framework conditions are not systematically treated, and this makes it difficult to evaluate the conclusions drawn from the studies. An analytical framework for assessing these theoretical underpinnings is developed below. It might properly be called a "meta-frame", as its only purpose is to give greater precision to the alternative interpretations of the cost per pupil indicator.

To develop this meta-frame, it would be appropriate to begin by examining the composite character of the indicator. However, the status of the indicator in a systems model of education cannot be directly identified because it is a ratio of a cost variable and a pupil variable. Therefore, the positioning of the numerator and the denominator in such a model must be investigated.

Interpretation of the complex variable "costs per pupil"

The "costs of education" and the "pupils" variables may have two positions in a systems approach to education and allow for four types of models for deriving interpretations of cost per pupil variances. The dichotomy in the "costs of education" reflects a distinction between a functional and an institutional definition of education. In the functional definition, costs represent the financial resources allocated for educational purposes. These costs thus act as an input (as opposed to output) indicator in a systems model of education. In the institutional definition, costs are associated with costs centres and reflect resources absorbed in production by these centres. Thus, costs may be considered to represent throughput characteristics rather than inputs allocated to education.

A similar distinction between input and throughput interpretations can be made with respect to the "pupils" variable. First, pupils may represent "raw material" entering education, and thus, as "inflows", numbers of pupils can be designated as input indicators. Second, pupils may be viewed as the clients of education. Expressed in units of "average daily attendance", numbers of pupils may properly be called throughput indicators. In this case, the numbers may be associated with activity levels in education, while in the first they certainly are not.

A two-way matrix for deriving the four alternative interpretations for the cost per pupil indicator can now be constructed. With respect to costs, a distinction can be made between input and throughput interpretations, determined by the functional and institutional definitions, respectively. With respect to pupils, an input interpretation can be given to inflow measurement and a throughput interpretation to any weighting procedure reflecting intensity of exposure to education. This matrix is shown in Figure 15.2.

In Figure 15.2, categories I and III may be associated with interpreting the cost per pupil indicator as a measure of "effort" and types II and IV with "efficiency", in accordance with the two approaches previously described. Adding the input/throughput distinction to these approaches gives:

- measurement of resources allocated (by society) to educational activities;
- measurement of resources used in educational activities;

Figure 15.2 **Interpretation scheme for the cost-per-pupil indicator**

		Pupils	
		Input	Throughput
Costs	Input	I	II
	Throughput	III	IV

- measurement of resources available to educational institutions;
- measurement of resources used in educational institutions.

The adequate interpretation of a cost per pupil indicator will depend on the type of variables used in its construction. Thus, the matrix can be employed to analyse whether the interpretation of the indicator is consistent with the costs and pupils data underlying its measurement, and the matrix is applied as a "consistency check" for the interpretation given to the indicator.

When an understanding of the indicator has been established, attention may be directed to the explanation of the observed variances. The four interpretations will focus on different implications of the variances. However, as any interpretation inevitably suggests specific connotations, the explanation will require careful analysis of the implications of the interpretation so as to avoid unwarranted conclusions.

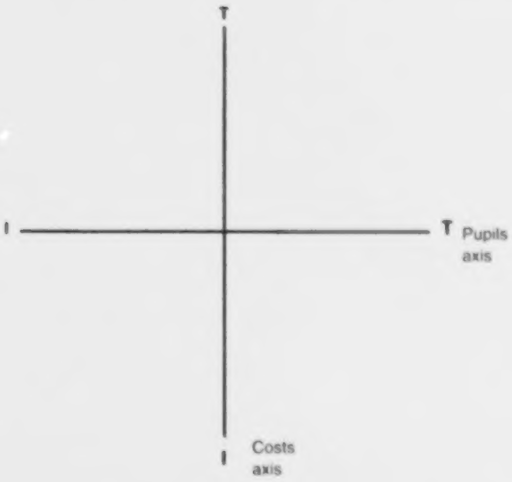
The first interpretation suggests that high costs per pupil reflect the high priority given to education by society. The second sees high costs as indicating resource-consuming production processes in educating pupils. The third hints at the idea that educational institutions could be considered relatively rich. The fourth suggests an unfavourable judgement on the costs of production in these institutions. Clearly, these connotations may be incorrect and the explanations suggested by them wrong. To ascertain the validity of the conclusions, further analysis of the variances will be needed. The next section considers how cost per pupil variances in each of the four interpretations may be examined.

Analysis of cost-per-pupil variances

In principle, the analyses would require developing complete system models for each interpretation. However, the focus here will be placed on the difference between the interpretations given to the cost and pupils variables, and the analysis will be limited to the resources and processes sectors in the systemic scheme presented in Figure 15.1.

From the two-way matrix developed above, a framework can be elaborated to help evaluate the judgements made on the observed indicator variances. This framework, shown in Figure 15.3, is presented as a co-ordinate system, in which one axis measures input and throughput characteristics with respect to pupils and the other input and throughput characteristics with respect to costs.

Figure 15.3 **Analytical framework: the co-ordinate system**



If the number of pupils is seen as an input variable, it is measured from the origin to the left, and as a throughput variable to the right. If costs are considered to be an input variable, then this variable is measured from the origin downward, and as a throughput variable from the origin upward.

The first step in analysing cost per pupil variances within the framework consists of identifying a "reference value" for the cost per pupil indicator in an intended comparison of education systems. This value may be derived from an education system considered "exemplary", or it may be an average value for the systems to be compared. This yardstick makes it possible to standardize measurement of costs and pupils on the axes chosen, depicted by the 45 degree line between them. Figure 15.4 gives an example for the input-output interpretation.

Scaling one of the two axes unequivocally produces the scaling of the other, so variances in actual costs can be depicted in the framework (Figure 15.5).

As a second step towards evaluating this variance, it is necessary to scale the input or throughput counterparts of the cost and pupil variables. A selection of relevant

Figure 15.4 **Analytical framework: the reference value**

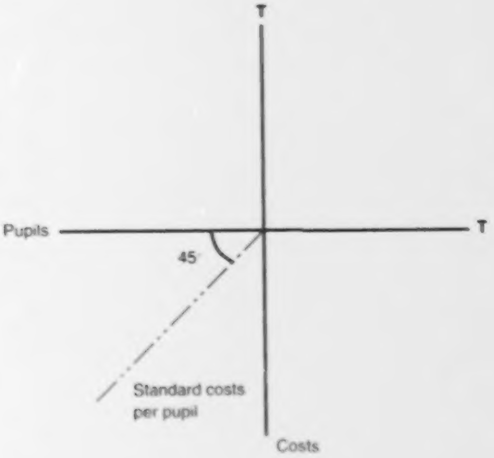
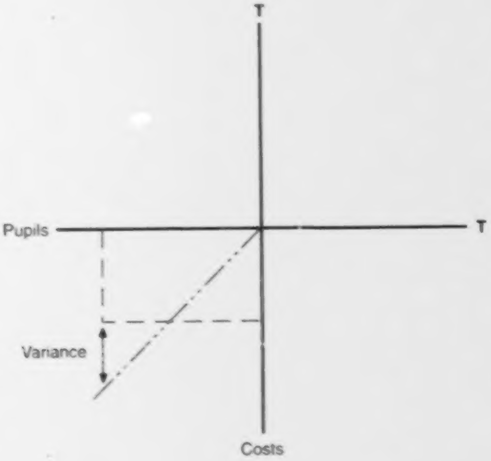

T
Pupils
T
45°
Standard costs per pupil
Costs

Figure 15.5 **Analytical framework: depicting variances in costs**


T
Pupils
T
Variance
Costs

variables, with specifications to be derived from the research literature discussed below, would include:

- for costs as an input variable: some "unit of teaching" as a throughput measure;
- for costs as a throughput variable: some "unit of teaching staff and materials" as an input measure;
- for pupils as an input variable: some "unit of knowledge transferred" as a throughput measure;
- for pupils as a throughput variable: some "unit of pupils' prior knowledge and abilities" as an input measure.

Given a specified set of variables, the conceptual framework can be completed by defining the quadrants left and measuring the units chosen. An example is presented in Figure 15.6.

Measurement may be standardized in the framework by starting from the reference value for costs per pupil. This reference value may be broken down into three components, since it is believed to represent standard costs of a standard quantity of units of knowledge transferred in a standard quantity of units of teaching to the pupils in the education system. These three standards may be depicted in the same way as the resulting "standard costs per pupil" by the 45 degree lines in their respective quadrants. Thereby, all axes will be fully scaled and the analysis of some observed cost per pupil variance may be presented as in Figure 15.7.

In Figure 15.7, the relative weights of the three factors explaining observed cost per pupil variances are indicated by the divergence of the ratios from the 45 degree lines. Figure 15.8 shows how the variance of the costs per pupil (indicated by angle z) is offset

Figure 15.6 **Analytical framework: identification of counterparts**

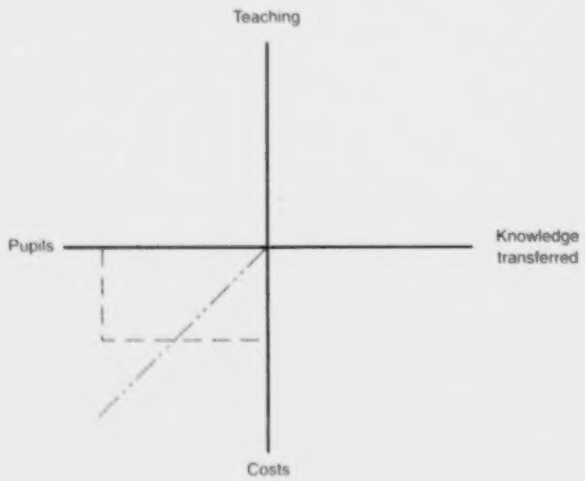
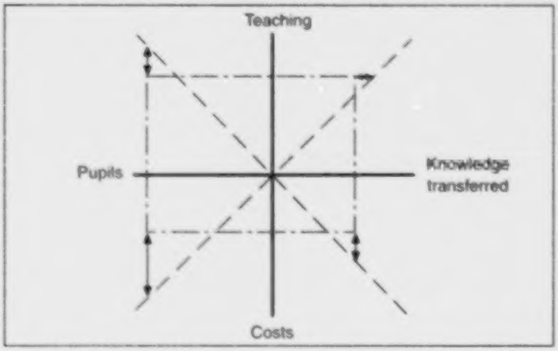


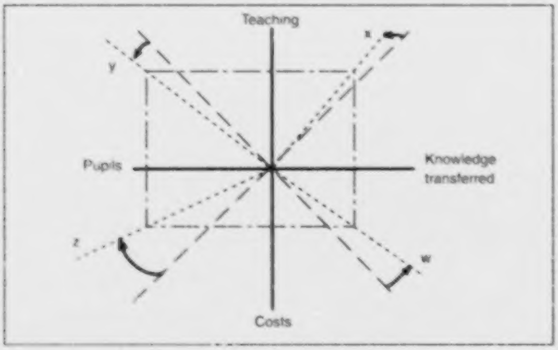
Figure 15.7 **Analytical framework: identification of variances**


Teaching
Pupils
Knowledge transferred
Costs

by the angles w, x, y. The angles w, x, y diverge in the opposite direction and all contribute to explaining angle z; z being equal to the sum of w, x, y, the relative importance of these factors in explaining the observed variance in the costs per pupil may be read from Figure 15.8 as $w > y > x$.

Thus, in this case, low costs per pupil might be explained in order of importance by: high "cost effectiveness" (measured in the fourth quadrant); low technical efficiency (measured in the second quadrant); and a relatively low quality of teaching of the system (measured in the first quadrant). As already mentioned, this explanation should be considered tentative, as long as the analysis is not embedded in a complete model of the

Figure 15.8 **Analytical framework: assessing explanatory weights**


Teaching
x
y
Pupils
Knowledge transferred
z
w
Costs

education systems to be compared. In particular, characteristics of the environment and the effects of education should be taken into account.

The above analysis can be presented in an algebraic equation:

$$C/P = C/K \times K/T \times T/P \qquad [1]$$

with C/K as the cost of knowledge transferred (the reciprocal of "cost effectiveness"), K/T as a measure of "quality of teaching" (knowledge transferred by a certain volume of teaching activities), and T/P as a measure of "technical efficiency" (level of teaching activities per pupil).

The other interpretations of the cost per pupil indicator described above give rise to similar analyses. For the input-throughput interpretation, where costs per pupil are supposed to measure resources used in educational activities, one may write:

$$C/P = C/D \times D/T \times T/P \qquad [2]$$

where D indicates deficiency levels in pupils' abilities and prior knowledge. Here C/D may be interpreted as an indicator of resources available to overcome deficiencies in education. The first quadrant again has an indicator for "quality of teaching" (in this case T/P), the second has an indicator for "technical efficiency" (D/T).

The analysis of the throughput-input interpretation, involving measurement of (physical) units of staff and materials (S) may use the formula:

$$E/P = C/K \times K/S \times S/P \qquad [3]$$

with C/K indicating "quality of teaching" (in the first quadrant), K/S indicating "(input) effectiveness" and S/P "available resources".

Finally, for the throughput-throughput interpretation one may write:

$$C/P = C/D \times D/S \times S/P \qquad [4]$$

with D/S representing physical resources allocated to educational assistance. It should be noted that C/S cannot be interpreted as the price of a unit of teaching staff and materials. With C indicating total institutional costs in the throughput interpretation and S indicating teaching staff and materials, price levels will be only partially reflected in the ratio. However, assuming that all expenditures are measured in purchasing power parity (PPP), as intended in the INES comparison of costs per pupil, variances due to diverging price levels may be considered eliminated in the indicator values.

The analytical framework presented above can be used as a guideline in the search for explanations of variances in costs per pupil. Starting from an interpretation of the cost per pupil indicator consistent with the definitions of costs and pupils used in the research, it is possible to judge to what extent the interpretation is consistent with the explanatory variables. To the extent that this set of variables is found wanting, explanations of observed variances in the research may be considered inconclusive. However, the framework should not be conceived as an "algorithm" for deriving the right explanation of cost per pupil variances. It is only intended to serve as a heuristic device to differentiate between the interpretations proposed in the literature and to examine the connotations and suggested meanings of these interpretations. As such, its goal is to create some order in the chaos of cost per pupil comparisons.

Examination of cost-per-pupil comparisons

The first cell in the matrix refers to the input-input interpretation. Walberg and Fowler (1987), as well as Eicher (1989), use a functional definition of the costs of education. The number of pupils is seen as a raw input in the educational process, which generates an increase in their knowledge. According to the matrix, therefore, Walberg and Fowler use an input-input interpretation which implies that costs per pupil measure resources allocated by society to educational activities. However, their interpretation of variances in costs per pupil is not very clear. On the one hand, costs per pupil are seen as an indicator of the share of schooling in the American economy, but, on the other, costs per pupil are interpreted in the context of a lack of incentives to improve educational productivity. According to the matrix, the proper interpretation would refer to the share of schooling in the American economy. The findings of Walberg and Fowler cannot be used to determine the causes of variances in educational productivity, as they propose (p. 13), as the interpretation is not consistent with the definition of pupils and costs used.

Since authors such as Walberg, Fowler and Eicher implicitly use the input-input interpretation, explanations for variances in costs per pupil have to account for "units of teaching" and "units of knowledge and abilities transferred". The analyses of Walberg and Fowler include the latter, since the authors use test scores and scores of the socio-economic status (SES) of the school district, which are assumed to be proxies for the "prior knowledge and abilities" of the pupils. As for the unit of teaching, the authors use average daily attendance rates (ADA), which combine the measurement of the number of pupils and the intensity of instructional contact (which is an operationalisation of units of teaching). Variances in costs per pupil, as an indicator of the share of schooling in the American economy, may thus be explained. Variances in costs per student as measured by Eicher, however, are very hard to interpret in terms of variances in national efforts, as his paper takes account neither of "units of teaching" nor of "units of knowledge transferred".

The second interpretation of costs per pupil, the input-throughput interpretation, is implied by Hirsch (1966), who combines a functional definition of education with a throughput measurement of the number of pupils. The study shows a certain inconsistency between data and interpretation. Contrary to the input-throughput interpretation implied by his research, Hirsch explains variances in costs per pupil in terms of the efforts of school districts dedicated to primary and secondary education and the efficiency of schools. He seems to base his explanation of these variances on a throughput-input interpretation on the one hand, and on a throughput-output interpretation of the indicator on the other.

Equation [2] represents the input-throughput interpretation implied in the data used by Hirsch. To use this equation to generate a valid explanation, a measure of prior knowledge and abilities, as well as a measure of some unit of teaching, have to be taken into account. Hirsch includes an index of scope and quality of education, in which some reference is made to a unit of teaching (by including the number of teachers per 100 pupils) and the number of college hours of an average teacher. Hirsch did not include, however, a variable for prior knowledge and abilities.

The third interpretation is determined by the use of an institutional definition of education and an input definition of the number of pupils. To this category belong the studies of cost per pupil variances performed by Hanushek (1989), Stern (1989), and

Sebold and Dato (1981). Hanushek concludes that the relation between costs per pupil and pupils' achievement is not consistent. Stern comes to a similar conclusion, stating that variances in costs per pupil cannot predict variances in pupils' achievement, while Sebold and Dato conclude that there is a positive relation between variances in costs per pupil and pupils' achievement. The explanation of observed variances in terms of pupils' achievements seems to be in line with the interpretation given to the cost per pupil indicator, *i.e.* as a measure of resources available to educational institutions. It may be analysed within our framework by application of equation [3].

Knowledge transferred and units of teaching staff and materials are variables that should somehow be included in an explanatory analysis. Stern and Sebold-Dato use socio-economic characteristics, size of school and population characteristics as additional variables in relating costs per pupil to pupils' achievement. Hanushek includes family backgrounds, initial achievement scores, characteristics of the peer group, and the influences of curriculum and instructional methods. Since in all three studies costs per pupil are related to test scores, "knowledge transferred" is (to some extent) included in the analyses. Units of teaching staff and materials, which provide an input measure of educational activities in schools, do not appear in the strings of variables used by these authors. However, Stern (1989) and Sebold and Dato (1981) do incorporate a unit of teaching, measured as daily student class hours, but their studies lack an input measure for education activity in schools. The results of these three studies relating variances in pupils' achievement to variances in resources available to schools must be used with caution since variances in pupils' achievements may be due to variances in educational inputs as well.

Most studies reviewed use a throughput-throughput interpretation of variances in costs per pupil: Hough (1981), Bowen (1981), Brinkman (1981), Riew (1966) and McLaughlin *et al.* (1980) all combine an institutional definition of costs with a throughput interpretation of the number of pupils. Following the two-way matrix, variances in costs per pupil should be interpreted as variances in resources consumed in educational institutions. Riew, Bowen, Brinkman, McLaughlin and Hough use interpretations which are consistent with the definitions of costs and pupils used. Hough states, however, that the variances in costs per pupil indicate that some schools consistently enjoy more favourable levels of expenditure than the mean of their group, an interpretation which would be consistent within the throughput-input approach.

Most of the studies mentioned try to interpret variances in costs per pupil as variances in efficiency of institutions. As a result, a whole range of alternative explanations and intervening variables are analysed: the size of school (Riew, McLaughlin, Bowen, Brinkman, Hough) and changes in size (Riew); teachers' salaries (Brinkman, Hough, Riew); location and condition of the building (Riew, Bowen); staffing ratio (Brinkman, Hough, McLaughlin); and type and complexity of curricula (McLaughlin, Bowen, Brinkman). Whether or not inclusion of these variables ensures the validity of the intended efficiency judgements is a difficult question to answer, but the analytical framework developed in this chapter can be of help. According to this framework, variances in costs per pupil can be decomposed in variances in three ratios, as in equation [4].

The S/P ratio, indicating the staff/pupil ratio, is used as an explanatory variable by Brinkman, Hough and McLaughlin. The ratio of physical resources allocated to support action (D/S) taken to improve the quality of the teaching staff is included in the studies of Brinkman, Riew and Hough. For the C/D variable, no operationalisation was found. It can therefore be concluded that only in the studies of Brinkman and Hough may (tentative) judgements of the efficiency of educational institutions be warranted. The findings of McLaughlin, Riew and especially Bowen have too many ambiguities in the explanation to obtain clear-cut efficiency judgements from them.

Conclusion

The analysis of the cost per pupil indicators to be presented within the INES system will be a very delicate affair, especially because a clear-cut interpretation of the indicators cannot be derived from the INES system itself. Neither the measurement of costs nor the measurement of pupils in the system will be free of ambiguity, as both allow for input and throughput characteristics of the data in OECD Member countries. However, if these ambiguities are clearly understood, possible biases in the explanations given of cost per pupil variances may be revealed and unjustified policy recommendations drawn from such explanations rejected.

Acknowledgements

The authors want to thank A.A. Spee for his comments on an earlier draft of this chapter.

Bibliography

BOWEN, H.R. (1981), *The Costs of Higher Education: How Much Do Colleges and Universities Spend per Student and How Much Should They Spend?*, Jossey-Bass Publishers, San Francisco, CA.

BRINKMAN, P.T. (1981), "Factors Affecting Instructional Costs at Major Research Universities", *Journal of Higher Education*, Vol. 52, pp. 265-79.

COLEMAN, J.S. *et al.*, (1966), *Equality of Educational Opportunity*, United States Department of Health, Education and Welfare, Government Printing Office, Washington, D.C.

EICHER, J.C. (1989), "The Financial Crisis and its Consequences in European Higher Education", paper presented at a conference of the Center for Higher Education Studies, University of Twente, Enschede, June.

HANUSHEK, E.A. (1989), "The Impact of Differential Expenditures of School Performance", *Educational Researcher*, Vol. 18 (4), pp. 45-51.

HIRSCH, W.Z. (1966), "Determinants of public education expenditures", *National Tax Journal*, Vol. 13, pp. 29-40.

HOUGH, J.R. (1981), *A Study of School Costs*, NFER/Nelson, Windsor.

McLAUGHLIN, G.W., MONTGOMERY, J.R., SMITH, A.V., MAHAN, B.T. and BROOMALL, L.W. (1980), "Size and Efficiency", *Research in Higher Education*, Vol. 12 (1), pp. 53-66.

NUTTALL, D. (1992), "The Functions and Limitations of International Education Indicators", in *The OECD International Education Indicators: A Framework for Analysis*, pp. 13-24, OECD/CERI, Paris.

RIEW, J. (1966) "Economies of scale in high school operation", *The Review of Economics and Statistics*, Vol. 48, pp. 280-287.

SEBOLD, F.D. and DATO, W. (1981), "School Funding and Student Achievement: An Empirical Analysis", *Public Finance Quarterly*, Vol. 9 (1), pp. 91-105.

STERN, D. (1989), "Educational Cost Factors and Student Achievement in Grades 3 and 6: Some New Evidence", *Economics of Education Review*, Vol. 8 (2), pp. 149-58.

van HERPEN, M. (1992), "Conceptual Models in Use for Education Indicators", in *The OECD International Education Indicators: A Framework for Analysis*, pp. 25-51, OECD/CERI, Paris.

WALBERG, H.J. and FOWLER Jr, W.J. (1987), "Expenditure and Size Efficiencies of Public School Districts", *Educational Researcher*, Vol. 16, pp. 5-13.

BEST COPY AVAILABL

Chapter 16

National Education Indicators and Traditions of Accountability

by

David K. Cohen** and **James P. Spillane
Michigan State University, East Lansing, United States

Education indicators are seen as a way to improve education by improving decision-making about education. Advocates argue that if policy-makers have better evidence on the performance of students, schools and teachers, then they will be able to make improved decisions about resource allocation and policy direction. The belief that scientific knowledge would improve political decisions is at least as old as Saint-Simon's dream of a social science, but the recent enthusiasm for indicators suggests a renaissance of this belief.

The ambitions are appealing, but can they be realised? In this chapter this question is discussed under several headings. One concerns the design of indicator systems. These systems could be created in many different ways, each with very different consequences for the information reported and the burden of collecting the data. A second set of issues concerns the fit between indicator systems and education, for the structure of school systems varies greatly among countries. A third set of issues concerns the ways in which indicator systems may interact with national arrangements for political responsibility, since these arrangements vary in ways that could affect the use of indicator data. A last set of issues concerns the influence that different traditions of professionalism may have on the design and operation of indicator systems.

Design of indicator systems

Systems of education indicators might be designed in any number of ways. They could focus chiefly on educational resources, such as the availability of books, teachers' qualifications and per-pupil expenditures. Or they could focus chiefly on educational outcomes, such as students' scores on various tests of achievement, school completion rates, admission to tertiary education, and job placements. Or indicator systems could be designed around the relations between resources and outcomes, such as differences in the "effectiveness" of schools at given resource levels.

These are great differences in the design of national indicator systems. Yet, social science offers little solid guidance about which emphasis would be preferable. Should indicators be oriented to educational inputs, or instead focus on instructional processes and outputs, including learning? The difference is crucial, for it would affect the knowledge that is produced about education, as well as affecting the burdens of constructing and operating indicator systems. For the sorts of data on resources that have considerable face validity with decision-makers and taxpayers, and that would be relatively easy to collect with reasonable validity – such as money spent on schools, or student attendance, or teacher qualifications – do not appear to be strongly related to the school outcomes that increasingly interest decision-makers (Hanushek, 1986). In contrast, resource data that have less face validity and are much more difficult to collect – such as teachers' knowledge of subjects, and how well they teach – seem to have a stronger relation to students' learning (Brophy, 1988; Smith and O'Day, 1990; Resnick and Resnick, 1989). One finds similar problems with evidence about student outcomes. Students' performance on standardized tests are relatively easy to collect, score and analyse. But increasingly it is argued that these tests measure only low-level skills, and would set the wrong sorts of targets for public policy and schools (Madaus, 1989, 1990; Linn, 1989; Fredrickson, 1984; Fredrickson and Collins, 1990). It is possible to design examinations that assess a much greater range of knowledge and skill, and that require more complex intellectual performances from students, but they are much more costly and complex to devise, administer and analyse.

A related design problem concerns the relationships between resources and outcomes. National indicator systems might report resources and results in isolated form, as part of an array of discrete indicators. Alternatively, they could focus on the connections, in an effort to draw attention to indicators of effectiveness. Those connections are a matter of considerable interest to researchers, and are increasingly of concern to decision-makers. The indicators movement itself can be viewed as an expression of that concern. But in addition to increasing dispute about how results should be conceived and assessed, there is also much uncertainty about several other key issues. What sorts of resources are most influential, for any sort of result? How do resources influence results? And how important is their influence relative to other factors? Researchers have learned a good deal about the relations between resources and outcomes during the past three decades, but they are far from having a well-specified "model" of schooling processes. Indicator systems that focused on the relationships between resources and results would be likely to draw attention to problems of validity.

Such differences could also affect the political consequences of indicator systems. An input-oriented system is less likely to stir controversy than an output-oriented system, simply because the latter will offer material for disputes over the effects of education and the effectiveness of schools. An indicator system that includes data on inputs, processes and outputs is even more likely to occasion political dispute, because it would focus attention on the effects of educational inputs and processes on outcomes.

Still another design issue concerns comparisons. A system that reports for a nation as a whole and compares it only with other nations is likely to be less burdensome to design and operate than one that also reports on intra-national comparisons among schools, localities, states, or regions. Such internal comparisons would raise difficult issues of sampling, units of analysis, reporting, data-collection burden, and cost. Internal comparisons also would be more likely to generate controversy. For they would fuel or stimulate disputes about equity in the allocation of resources, or differences in effectiveness, or both. Indicator systems that also included comparisons of ethnic groups or social classes would offer even more occasions for dispute.

These design issues interact. A system that included only inputs and international comparisons would be much less complex and burdensome than one that included inputs, processes and outputs, and compared within nations across states or regions, schools, classes and social or ethnic groups. The more complex indicator systems are, the more they fascinate researchers and many decision-makers, but the more opportunities for controversy they open up. The simpler they are, the less interesting they are to researchers, but the fewer occasions for controversy they would generate.

Several other design issues are less technical. One critical matter is whether indicator systems are arranged chiefly to lead policy and practice or to monitor it in current terms. For example, measures of educational results could be designed so as to lead change, if they included "authentic assessments" – *i.e.* open-ended history questions, or math problems that required written explanation and justification of answers. Measures of instructional processes could be designed to assess teachers' knowledge of subject matter, or to observe teachers at work. Alternatively, measures of output and process could be designed to follow current practice. Students' learning could be assessed in traditional terms with standardized, multiple choice tests. And teachers could be asked to report the texts they used, the topics they covered, and how much time they spent on each. An indicator system that followed current practice would be less complex and costly to design and operate than a system that was designed to lead practice and policy. A system that followed current practice could be controversial, depending on its design and political circumstances. In contrast, a system that was designed to lead practice probably would produce more controversy, since its reports would pose many questions about the adequacy of instruction, resources and policy. That might threaten the indicator system and the seriousness with which it was taken. But such a system also might be more likely to move opinion and education towards better instruction.

These design issues cannot be resolved here. National tastes in education vary, as do researchers' views about these matters. But for the purposes of the following discussion a system is proposed that seems most appropriate to the state of knowledge about education. This indicator system has several key attributes:

i) It is defined and organised as a means of reporting on status and change in education systems. It is not defined and organised as a vehicle for doing other research on the operation of those systems or their sub-units.

ii) It monitors a range of educational resources, from the most rudimentary (such as unit expenditures) to the more complex (such as teachers' knowledge of the subjects they teach).

iii) It monitors traditional measures of educational results, such as school completion, but also the best available standardized measures of students' performance in fairly complex intellectual tasks, such as those used in some parts of the United States National Assessment of Educational Progress (NAEP). When paper-and-pencil measures of more complex intellectual tasks are developed, some should be included in the set of outcome indicators.

iv) It is designed not to permit the analysis of interrelationships between resource inputs and results.

Some commentators in this volume view such an indicator system as analagous to the array of "dashboard lights" in an automobile. The dashboard indicators offer information on the operating status of the vehicle. But they offer little or no insight either into the nature of the systems being monitored or into possible pathologies therein. Drivers can tell if there may be a problem, but with few exceptions (petrol empty), more detailed investigation, beyond the scope and substance of the indicator system, would be required to diagnose problems or devise solutions.

The structure of school systems

Many advocates seem to assume that national sets of education indicators would be helpful in any nation's school system. But education systems differ markedly across societies. For example, some national school systems are highly centralised, as in France or Singapore, while others are decentralised. Some national systems contain strong elements of both central and local control, as in the United Kingdom. Some countries, moreover, lack a national school system, as in the United States and Australia. Instead, control over schools is spread across a complex federal structure of government.

Would a national system of educational accounts be equally suitable in all such systems? Can one assume an equally good fit between indicator systems and school systems?

At first glance, national indicator systems would seem to make more sense in the highly centralised OECD countries, because in such cases, the policy is concentrated. National indicators would appeal to the salient policy-making audiences at the same time as they speak to the policy more generally. In France or Singapore, an indicator system would occur in a polity for which the data would be valid. And it would be sponsored by a government for which the data might be salient.

Matters are somewhat different in the more decentralised countries. Consider, for instance, such federal governments as those in Canada, Australia or the United States. In these countries, indicators seem to pose a political paradox. On the one hand, the multi-level government structure seems to argue for a multi-level indicator system, since many key decisions are made at a level other than the national one. A national indicator system would only bear on state and local decisions if the polity and society were homogenous, and there were no variation among jurisdictions. Yet many societies are anything but homogeneous, and the variations among jurisdictions are great. The problem, however, is that designing and operating a multi-level indicator system is difficult. It would be a daunting intellectual task to devise a valid system for states or provinces as well as the nation as a whole. Such a system would add greatly to costs and technical problems. Additionally, there would be many political difficulties associated with a multi-level system. For instance, decisions about system design and content, management, analysis, and reporting all would have to be made at two or three levels of government.

Centralisation and decentralisation are not, however, stable attributes of political systems. While interest in national indicator systems has been increasing during the last decade, so has interest in decentralising school systems. This is to be expected in the United States, where decentralisation is a lively political tradition. But several formerly highly centralised Australian states also have tried to devolve authority to schools or regional offices. And France and Singapore, both highly centralised states, are moving towards some devolution of authority and operations in education.

Competing policy directions are nothing new to education. But what can be made of this curious juxtaposition? Is education heading in two directions at once? Are national indicator systems only a throwback to older ideas about administration and service organisation? Would indicators work in decentralised education systems?

The answer is that they could, but probably not in the same way as they might work in highly centralised systems. A national system of educational accounts with purely national reporting might be seen as a service to the local operating units in a decentralised system, providing those units with broad evidence on system performance. But such comparisons would have limited value, since in a truly national system, sub-units would have no indicator data on their own performance. If state, regional and local units were to be able to compare themselves realistically with the national picture, then a multi-level system would be required.

It may be noted that indicators have more than instrumental value – they are also a potent political symbol. For instance, recent governments in Great Britain, Australia, the United States and Canada all have been sceptical of central power, and have tried to disentangle the state from a variety of social and economic interventions previously adopted. Yet political developments in education and their economies have impelled these governments to take an activist stance on some educational issues. Rather than trying to diminish central power, governments in Great Britain, Australia and the United States have been expanding central influence in education, partly because educational improvement is seen as a way to improve the economy. Yet all three governments face growing fiscal problems. They believe that there is little money for expensive new policy initiatives, and several have ideologies that might make such initiatives seem incongruous, even if there were more money. In such circumstances indicators are useful. They appear to be a bold reform initiative, partly because they assert a larger role for the central government and partly because they embody a scientific stance in policy-making (Salter and Tapper, 1985). Yet indicators also are quite inexpensive as central policy initiatives go. What is more, they commit a government to no substantive policy position, but only to effectiveness and efficiency in schools, and to basing government policy on the best evidence. Indicators also can help to position central governments – or for that matter states or provinces – as an important contributor to policy debates and perhaps policy-making.

Indicators also may offer central governments a way to influence political discourse without the burdens of policy. For governments need only establish and operate an indicator system, and report its results, to stand a good chance of gaining a larger voice in the policy discourse. Indicators may be a way for central government to "do something" about education while avoiding the costs of devising or operating policies.

Finally, the interest in indicators may signal some changes in thinking about education. There is a growing or perhaps renewed awareness that schooling is in many ways a cottage industry. Educators and policy-makers in many nations have been discovering or rediscovering the crucial importance of local initiative, intelligence, and knowledge in schools. They also have been recognising the enormous variations among schools, even within seemingly homogeneous communities. Yet these discoveries have coincided with growing pressures for state or national action to renew education, pressures that are perhaps greater than ever before. This is a time in which problems are great but money is limited. It is a period in which action by government seems essential, but in which relationships among central, local and other government agencies seem unusually fluid.

Indicators may offer a way for governments to respond to these circumstances without deep policy entanglements. Perhaps indicators can help to frame new relationships, in which central agencies have more influence in educational matters, but in which local educators have more opportunities to act intelligently and constructively. No one knows exactly what such arrangements might be, but several nations seem to be trying to invent them. Indicators may be part of a new portfolio for central government in education, signalling impartial oversight, standard setting, and central influence and responsibility. It remains to be seen whether governments that take such steps can avoid being drawn into more central control or regulation of education.

Traditions of political responsibility

In all the discussions of national indicator systems there has been little attention to nations. Yet government and politics vary greatly among them. Some developed nations are stable democracies while others are less stable and less democratic. Some are vast and others tiny. Some are wracked with civil strife but others, that once were political cauldrons, now are so stable as to make civic discord difficult to imagine.

These are great differences, that could powerfully affect the design and use of social and education indicators. Even if the activity is confined to the OECD countries, remarkable differences appear. The most salient of these is the arrangement of political responsibility. Some countries have relatively clear lines of political responsibility and tight arrangements for holding governments accountable, while the situation is less clear in others. Such variations may affect both the design and use of indicator systems, since political information needs and uses often differ between tightly accountable and more diffuse systems.

What distinguishes the two sorts of democracies? Research on government, politics, and policy-making suggests several considerations. Some democracies seem more coherent, for they centre political responsibility in a parliament. In contrast, others diffuse responsibility among various elements of government. Some democracies also focus responsibility more clearly because they have relatively few political parties that are strongly disciplined, and a modest array of relatively stable national interest groups. Others have more parties, or they are weakly disciplined. No political systems are pure types, and the arrangements for political responsibility do not vary smoothly across nations. For instance, some parliamentary governments have stable and well-established interest group politics, but a large array of weakly disciplined political parties. Nonetheless, some systems are closer to one end of the continuum or the other. The United States is almost certainly the best example of diffuse responsibility among the industrial democracies. And Great Britain may be the leading example of more coherent arrangements for political responsibility. Most industrial democracies are somewhere between these extremes. A few of these differences are outlined below.

The coherence of central governments is one salient element in the structure of political accountability. Parliamentary governments often are thought to have exemplary arrangements for democratic responsibility. For the national executive arises from parliament rather than being separately elected, and ministries are directly responsible to parliament (Almond and Powell, 1978; Smith, 1988; Olson, 1980). The relationship between legislative and executive branches of government therefore is relatively tight, and the sources of government action relatively clear. Voters thus are thought to have a less difficult time figuring out what actions to hold government officials responsible for. The judiciary has a relatively minor role in social policy (Smith, 1988; Kogan, 1986).

That also contributes to coherence within central government, and to clarity in attributions of political responsibility. The British government often is held up as a good example, but parliamentary governments are found throughout the OECD membership.

The United States' central government is often regarded as another example. Here the legislative and executive branches are separately elected, and their powers carefully divided. Indeed, central power is parcelled out among judicial, executive and legislative branches of the central government (Almond and Powell, 1978; Olson, 1980; Lundy, 1989), a pattern that is repeated in the states. The relationship between executive and legislative branches is relatively loose, and the sources of government actions may therefore be unclear. Voters are thought to have a difficult time figuring out what actions to hold government officials responsible for. The judiciary's large role in social policy (Levin, 1987; Yudoff, 1979; Hague and Harrop, 1987; Blondel, 1972) further diffuses responsibility for government actions. The United States government often has been held up as a model of balance and restraint, but few other governments in the developed world equal its fragmentation of formal responsibility.

Party and interest group politics are two other important influences on arrangements for political accountability. The more diffuse party and interest group politics are, the more difficult it can be to attribute responsibility for official action. Some countries, such as the United Kingdom, have a few fairly stable and disciplined political parties, while others, such as Italy or Israel, have more parties and more parliamentary instability. Some research studies, however, indicate a decline in party discipline in the United Kingdom during the 1970s (*e.g.*, Epstein, 1980; Schwartz, 1980). Yet, when compared to North American political parties, British political parties enjoy considerable cohesion and are well disciplined. In some cases, party discipline also is weaker. In systems with many parties, coalition governments are common, and political alignments can be more fluid. Both can obscure the sources of government action, and make it more difficult for voters to attribute responsibility for policy correctly. Sometimes there is little responsibility to attribute, because there is little action. Some coalitions are so unstable that clear policies would cause the government to come unglued. In contrast, in systems with fewer and more disciplined parties, political alignments are more clear, and it is easier for voters to attribute political responsibility. Party discipline supports executive dominance (Smith, 1988; Rose, 1980). In such cases government is more likely to speak with one voice.

Some parliamentary systems also have relatively stable alignments of interest groups, and a modest array of such groups. Great Britain and France, for instance, have had fairly stable networks of interest groups at the national level. The number of groups is relatively small, and they have had well-established relations with government and other interests. Divisions over issues and ideology, though often deep, have been few in number and stable over time (Grant and Richardson, 1982). The relative stability of national organisations and relationships has made it difficult for new interest groups to organise and become influential. This situation has contributed to the relative stability of governing majorities, and to leading parties' capacity each to speak with one voice on matters of policy.

In contrast, party politics in the United States seems undisciplined. There are only two major parties, but there are deep ideological divisions within each. There is, for example, a remarkably high level of disagreement about educational problems and policies within each party. Party discipline always has been relatively weak in comparison with parliamentary systems (Epstein, 1980), and has grown weaker in recent years (Collie, 1985). Based on a review of the relevant literature, Collie (1985) notes that the

BEST COPY AVAILABLE

results indicate an erratic but overall decline in party discipline and interparty conflict in the United States. Additionally, the differences between Congressional and Presidential parties and politics have increased in recent decades, even further weakening party discipline across branches of government. As a result, central government action regularly depends on legislative coalitions that cross party lines (Olson, 1980; Lowenberg and Patterson, 1979), and policies tend to blur party identification. Compromise often is said to be a genius of American government. But one consequence is that voters have a difficult time assigning responsibility for government action.

Additionally, United States' national politics has seen a large and still growing array of interest groups. Though some have had fairly stable relationships in and around government, the division of formal responsibility and the lack of party discipline have encouraged a pattern of weak and shifting coalitions among interest groups, and between them and government. The divisions of political responsibilities, the general weakness of government, and the very large part that private agencies have played in public policy all have created many opportunities for groups to become organised. The arrangement is fluid and encourages relatively permeable government. These features have become even more pronounced as single issue politics has become more common, and as social policy has expanded. The population of interest groups in education, for instance, is much greater now than it was in 1960. And in both national and state governments, previously settled relationships in education politics have become unsettled.

These differences could well affect the use of indicator data, because governments with clear lines of responsibility place a considerable value on consistency in matters of policy. Governing and opposition parties would each be likely to speak with one voice in interpreting evidence from indicators, as in many other matters. The chances of consistency would be even stronger in nations with disciplined parties and stable interest group politics. In contrast, evidence from indicators probably would be perceived in a less discliplined way in systems with more diffuse government, less disciplined parties and/or more diffuse interest group politics.

But disciplined parties, coherent government, and stable interest group politics are not an unalloyed good, since in focusing and disciplining the use of indicator data, they might also constrain it. The very structures that clarify political responsibility also can insulate governments from disturbing or dissonant evidence about schools' performance. It is just such evidence that indicator systems might turn up. Once a governing majority is formed in a parliamentary system, especially one with highly disciplined parties, the government can be quite well insulated until the next election, barring some great political shock. Such evidence can be endlessly considered in parliamentary debates, all to no avail in policy. The Thatcher government's policies on education are a case in point, as are those of the Greiner government in New South Wales, Australia. Both educational policies have been controversial, but the main lines of policy have been carried out nonetheless.

In contrast, governments marked by weak party discipline and a larger number of parties are less likely to develop a common understanding of evidence from indicator systems. Such governments may suffer from multiple and chronic conflicts in political perception, but they might also more effectively consider evidence from indicators, simply because their internal divisions and weaker party structure open them to a greater range of information.

The United States government is a good example. Agreement on the evidence produced by an indicator system would be unusual in this country, partly because government was carefully designed to reduce consistency at the centre. For instance, though the recent education summit by the President and governors was a very unusual step, it took only a few months for divisions between legislators and the White House to emerge. This dissonance can be seen as a barrier to effective government, but there is another way to look at it. The United States government is remarkably various and permeable. If there were a national system of education indicators, many executive and legislative agencies and congressional committees might be eager consumers. The remarkable diversity and energy of interest groups also would increase the chances that indicator data would be put to many different partisan uses, and thus would have many potential consumers outside government. America might offer indicator data many points of entry to national policy debates, precisely because its government is so fragmented and its interest groups so active.

Is coherence in government an asset or a liability in the uses of evidence from education indicators? It depends on how one chooses to assess information use. If highly focused use that is tied closely to strong party government is deemed desirable, then coherence would be an asset. If broader but less focused use seems appropriate, then coherence in government may be less of an advantage. In this matter it is easier to delineate patterns of social action than to evaluate their merits.

Professionalism and political culture

Whatever part education indicators turn out to play in democratic accountability, formal governance will not be the only influence at work. Various dynamic features of political and education systems also are likely to shape the uses of indicators. Professionalism is such a feature, and it merits particular scrutiny. One reason is that professionals will be among the chief interpreters of such data. Another is that professionalism varies across countries and school systems.

Professionalism and public service

Democracies differ in the extent to which the national civil service is independent of political governance. The French and British government services are well known for such independence, and the United States' government service is equally known for its relative lack of independence (Hague and Harrop, 1987; Steel, 1979). Specialists in public administration often have argued that independence has produced a professional and higher quality administration in France and Great Britain than in the United States.

Education systems also differ in the extent to which they are independent from politics, and thus perhaps also in the professionalism that education officials can cultivate. Educators in most European school systems traditionally have been relatively independent of national politics, even though those systems were directed by national ministries. That tradition also was well developed in many British Commonwealth school systems. By contrast, public education in the United States has been a creature of local politics, and schools in many localities have been deeply entangled in politics.

It seems reasonable to think that more professional and independent government services would be more receptive to evidence from scientific sources. One reason is that professionalism depends on specialised expert knowledge. Certainly many professionals in public administration have been educated in conformity with that view. Another reason is that civil servants are increasingly expected to have a solid grounding in the social sciences. Indeed, social science is becoming the dominant language of discourse for advanced education in public administration. Still another reason is that the culture of government agencies can affect their stance towards advice and new knowledge. American social scientists have distinguished local governments that adopt a universalistic, "good government" posture from those that adopt a more particularistic and traditional approach to politics (Banfield and Wilson, 1963; Boyd, 1976). Particularistic governments are said to be less concerned with substantive policy decisions, less likely to seek a variety of advice about decisions, and more concerned about patronage or building contracts. Universalistic governments may well be more concerned with substantive policy decisions, more likely to seek a variety of advice about decisions, and less concerned about patronage. Governments that embody a more universalistic political culture, in which independence, expertise, and professionalism figure prominently, are likely to encourage wider consultation and to give more attention to professional advice. Agencies that embody a more particularistic culture in which patronage and connections figure prominently would be less likely either to encourage consulation or to attend to professional advice.

From this angle it seems plausible to think that more autonomous education systems or national civil services would be more receptive to evidence from education indicators. Hence one might expect the British or French civil services to offer a more ready audience for indicator systems than that of the United States.

But independence in school systems is not a stable quality, nor is professionalism the only source of interest in scientific information. In many nations education is rapidly being drawn into national politics, and is increasingly a focus for controversy. In the United States, education has been steadily more politicised since 1945. It also has been increasingly drawn into national politics in Great Britain. And in New South Wales, the politicisation of education has been more recent, though no less vivid. In each of these cases, political pressures have somewhat eroded the independence of school systems. But the increasing politicisation of education also has been accompanied by a growing appetite for evidence of system performance. The growing interest in indicators seems to be at least in part a consequence of this development.

Craft knowledge establishments

Education systems differ in the extent to which they cultivate professional craft knowledge. Some rely heavily on institutions that create, preserve and apply craft knowledge of teaching and administration. In contrast, others have no such institution, and rely on bureaucratic means to collect and diffuse knowledge about practice. The French and British have inspectorates. Inspectors are selected for their demonstrated knowledge of teaching and administrative practice, and for their proven capacity in practitioner roles, rather than on university degrees or mere seniority. School inspectors play a key role in decisions about professional promotion in many of these systems. They have had a key role in decisions about professional advancement for teachers and administrators, for instance. Such advancement depends on performance in practice, and demonstrated knowledge of practice, under the scrutiny of inspectors. Teachers can neither advance in service nor move into administrative posts without passing several inspections of their practice and their knowledge of practice.

Inspectorates also have sometimes taken a large role in promoting educational improvement. But whatever their particular assignments, the inspectorate is an agency that deliberately cultivates craft knowledge of educational practice. The chief basis for the inspectors' work is their own knowledge of practice. Hence the inspectorates become repositories of craft knowledge of teaching, administration and school improvement.

In contrast, the United States and many other systems treat decisions about professional advancement and school improvement in bureaucratic fashion. Craft knowledge is not institutionalised, and there is no agency that could make it widely available for use or circulation within educational agencies. Decisions about educational improvement and professional advancement are also made bureaucratically. Teacher promotions are based on years in service, the approval of administrators, and on advanced degrees in universities. None of these bear much relation to practice or knowledge of practice. Additionally, school improvement is managed by central administrators who work at some distance from instruction.

Would inspected school systems be less amenable to evidence from indicators? The inspectorate does embody a very particular conception of knowledge, and a very particular way of bringing it to bear on educational practice. If inspection works, it is because inspectors are connoisseurs (Eisner, 1985) of pedagogy and school management, and because they are authorised to use that knowledge in significant decisions (Kogan, 1986). Inspectors have accumulated knowledge through long service in practice. Much of what they know is tacit, specific to situations, and hence quite particular and personal. Their authority derives from experience, from passing inspections themselves, and because they inhabit a role that enables them to bring craft knowledge to bear on key decisions about educational practice.

Inspectors might resist evidence from indicators, since social science could seem a departure from craft knowledge, and perhaps a threat to it, since social indicators are an effort to apply the self-conscious knowledge of social science to education. The knowledge is articulate and depends little on educational experience. Such knowledge is quite remote from connoiseurship. And the authority of social science derives from its presumed independence, impersonality, and objectivity, rather than from close familiarity with educational practice. Perhaps such systems would be more open to evidence from indicators, since they lack much of an alternative knowledge base for decision-making. But these education systems seem to have no effective mechanisms for bringing social science knowledge to bear on practice.

Inspectorates have their share of disadvantages, too. They have been closed communities in some respects, and inspected systems have not been quick to use social science research (although there is some evidence that the Inspectorates in Great Britain have moved to use social science; Lawton and Gordon, 1987). But systems that have no craft knowledge establishment, and instead rely on social research for practical advice, have enormous problems of their own. For instance, social scientists in the United States recently "discovered" that school improvement works better when well informed and experienced staff are on the spot, and can offer continuing advice and assistance to practitioners. Such advice and assistance are of course the inspector's stock in trade. That this was seen as a discovery speaks volumes about the institutionalised ignorance that the absence of a craft knowledge establishment creates.

All of this suggests that education indicators might find a more ready audience in countries such as the United States, that have no institutionalised craft knowledge establishment, than in nations like France or Singapore, in which the inspectorate still is a key agency. But if that is encouraging for the adoption of indicator systems, it is not encouraging for their impact on practice. For systems that are not inspected have very weak arrangements for guiding and supporting change in practice.

Indicators and knowledge use

Finally, there are important national differences in the capacity for using social science, and in the disposition to use it. Some governments have a well-established tradition of using social science, and well-developed capacities for doing so as well. Sweden and the Netherlands are good cases in point. Other governments have less well-established traditions of use, and less capacity. Great Britain and Australia are examples here. The United States' government has added considerable capacity to use social research since the early 1960s, but there is less evidence that traditions of using social science have taken deep root. It seems reasonable to suppose that governments with weaker traditions and lesser capacities will be less enthusiastic about adopting or using evidence from indicators, or less deft at making use of it, or both. It remains to be seen whether indicator-like approaches that expand the knowledge base available to governments will thereby also expand the capacities and dispositions to use the information.

It seems more likely that variations in the use of evidence from indicators would depend on other influences – such as the esteem of science, the education of members of the government service, and the national experience with social science reporting on education. Some countries have a wealth of such experience, while others have relatively little. Some nations have little experience with internal social science comparisons of educational resources and results, while others are habituated to such comparisons.

The strength of scientific and professional elites also is likely to affect the demand for education indicators. Many members of such elites strongly identify with scientific knowledge, through their own education as well as their employment and social status. These elites' increasingly important role in and around government already has increased the scientific information available, and thus added to the problems of information overload in government. They are an important potential constituency for indicator systems, and these have considerable symbolic importance for them. National differences in the size, influence, or political mobilisation of such elites – in academies, through legislatures, in executive agencies, and the like – seem likely to influence receptiveness to and use of indicator systems.

Conclusion

Throughout the text, the use of evidence from indicators has been discussed without taking into account the nature or meaning of use. Some advocates of indicator systems seem to believe that social science evidence deserves to be used because it is more authoritative for decision-making – *i.e.* social science is considered to be objective, unlike professional judgement or popular opinion. These advocates also appear to assume that if scientific evidence on education is made available to policy-makers, they will use it, that is, base decisions on it. Even putting aside difficulties in these claims of objectivity, there is a further problem: Social science appears to be only dependently authoritative. That is, policy-makers are unlikely to take it seriously unless it accords with their pre-existing beliefs and values (Weiss and Bucuvalas, 1980; Lindblom and Cohen, 1977). Decision-makers of all sorts screen new information for evidence that it fits with established beliefs and practices. Evidence that does not fit tends to be screened out, or re-framed so that it does (Simon, 1960; Braybroke and Lindblom, 1963; Weiss, 1977; Lindblom and Cohen, 1977; Janis, 1989). This accords with an appreciation of the limits on human rationality, one of the great themes in modern social science.

This does not mean that evidence from indicators would be ignored. Professional public servants probably are more likely to read reports on education indicators than patronage appointees. But professionals may be no more likely to attend to evidence that is inconsistent with their established beliefs and practices. Several commentators in fact argue that it is political or economic incentives rather than disinterested analysis that stimulates civil servants and educators to make use of evidence from indicator systems (Lindblom and Cohen, 1977).

One possible consequence is that education indicators might be widely adopted, but not much used in decision-making. Advocates see indicators as a way to improve policy-making by improving what policy-makers know, but researchers who study the uses of social science find that it rarely has a direct effect, either on decisions or knowledge. At best, research contributes to policy-making through broad and diffuse processes of "enlightenment", *i.e.* by affecting climates of opinion (Weiss and Bucuvalas, 1980; Cohen and Garet, 1973; Lindblom and Cohen, 1977; Nisbet and Broadfoot, 1980). If so, indicator systems are likely to exert much more influence on opinion than on decisions. And whatever effect indicators have on knowledge or decisions is likely to be indirect, mediated through existing beliefs and values rather than supplanting them. Though these inferences offer some encouragement about the use of indicators, they offer little encouragement about accurate or objective use.

Bibliography

ALMOND, G.A. and POWELL, G.B. (1978), *Comparative Politics*, Little, Brown and Company, Boston, MA.

BANFIELD, E. and WILSON, J. (1963), *City Politics*, Harvard, Cambridge, MA.

BLONDEL, J. (1972), *Comparing Political Systems*, Praeger Publishers, New York.

BOYD, W. (1976), *Community Status and Conflict in Suburban School Politics*, Sage Publications, Beverly Hills, CA.

BRAYBROKE, D. and LINDBLOM, C. (1963), *A Strategy of Decision: Policy Evaluation as a Social Process*, The Free Press, New York.

BROPHY, J. (1988), “Research on Teacher Effects: Uses and Abuses”, *Elementary School Journal*, Vol. 89 (1), pp. 3-21.

COHEN, D. and GARET, M. (1973), “Reforming Educational Policy With Applied Social Research”, *Harvard Educational Review*, Vol. 45 (1), pp. 17-43.

COLLIE, M.P. (1985), “Voting Behavior in Legislatures”, in Lowenberg, G., Patterson, S. and Jewell, M. (eds.), *Handbook of Legislative Research*, Harvard University Press, Cambridge, MA.

EISNER, E.W. (1985), *The Educational Imagination: On the Design and Evaluation of School Programs*, Macmillan, New York.

EPSTEIN, L. (1980), *Political Parties in Western Democracies*, Transaction Books, New Brunswick, NJ.

FREDRICKSON, N. (1984), “The Real Test Bias: Influences of Testing on Teaching and Learning”, *American Psychologist*, Vol. 39, pp. 193-202.

FREDRICKSON, J. and COLLINS, A. (1990), “A Systems Approach To Educational Testing”, *Educational Researcher*, Vol. 18 (9), pp. 27-32.

GRANT, J. and RICHARDSON, J. (1982), “The British Policy Style or the Logic of Negotiation?”, in Richardson, J. (ed.), *Policy Styles in Western Europe*, George Allen and Unwin, London.

HAGUE, R. and HARROP, M. (1987), *Comparative Government and Politics: An Introduction*, Macmillan, London.

HANUSHEK, E. (1986), “The Economics of Schooling”, *Journal of Economic Literature*, Vol. 24 (3), pp. 1141-1177.

JANIS, I.L. (1989), *Crucial Decisions: Leadership in Policy-making and Crisis Management*, Free Press, New York.

JOHNSON, S.M. (1990), *Teachers at Work: Achieving Success in our Schools*, Basic Books, New York.

KOGAN, M. (1986), *Education Accountability: An Analytic Overview*, Hutchinson and Co. Ltd, London.

LAWTON, D. and GORDON, P. (1987), *The HMI*, Routledge, London.

LEVIN, B. (1987), "The Courts as Educational Policy-makers in the USA", in Boyd, W. and Smart, D. (eds.), *Educational Policy in Australia and America: A Comparative Perspective*, pp. 100-128, Falmer Press, New York.

LINN, R.L. (1989), "Accountability: The Comparison of Educational Systems and the Quality of Test Results", Vice Presidential Address for the Measurement and Research Methodology Division of the American Educational Research Association, San Francisco, CA, April.

LINDBLOM, C. and COHEN, D. (1977), *Usable Knowledge: Social Science and Social Problem Solving*, Yale University Press, New Haven, CT.

LOWENBERG, G. and PATTERSON, S. (1979), *Comparing Legislatures*, Little, Brown and Company, Boston, MA.

LUNDY, P. (1989), *Parliaments in the Modern World*, Gower Publishing Company, Brookfield, Vermont.

MADAUS, G.F. (1989), "The Distortion of Teaching and Testing: High-Stakes Testing and Instruction" (unpublished paper), Boston College, MA, June.

MADAUS, G.F. (1990), "The Influence of Testing on the Curriculum" (unpublished paper), Boston College, MA, November.

NISBET, J. and BROADFOOT, P. (1980), *The Impact of Research on Policy Practice in Education*, Aberdeen University Press, Aberdeen.

OLSON, D.M. (1980), *The Legislative Process: A Comparative Approach*, Harper and Row Publishers, New York.

RESNICK, L.B. and RESNICK, D.P. (1989), *Assessing the Thinking Curriculum: New Tools for Educational Reform*, Learning Research and Development Center, Pittsburgh, PA.

ROSE, R. (1980), *Do Parties Make a Difference?*, Chatham House, Chatham, NJ.

SALTER, B. and TAPPER, T. (1985), "The DES", in McNay, I. and Ozga, J. (eds.), *Policy-making In Education: The Breakdown of Consensus*, Pergamon Press, Oxford.

SCHWARTZ, J.E. (1980), "Exploring a New Role in Policy Making: The British House of Commons in the 1970s", *American Political Science Review*, Vol. 74, pp. 22-37.

SIMON, H.A. (1960), *The New Science of Management Decision*, Harper and Row, New York.

SMITH, A.T. (1975), *The Comparative Policy Process*, American Bibliographical Center and Clio Press, Inc., Santa Barbara, CA.

SMITH, B. (1988), "The United Kingdom", in Rowat, D.C. (ed.), *Public Administration in Developed Democracies: A comparative Study*, Marcel Dekker, Inc., New York.

SMITH, M. and O'DAY, J. (1991), "Educational Equality: 1966 and Now", in Verstegen, D. and Ward, J. (eds.), *Spheres of Justice in Education: The 1990 American Education Finance Association, Yearbook*, pp. 53-100, Harper Business, New York.

STEEL, D.R. (1979), "Britain", in Ridley, F.F. (ed.), *Government and Administration in Western Europe*, St. Martin's Press, New York.

WEISS, C. (ed.) (1977), *Using Social Research in Public Policy Making*, Lexington Books, Lexington, MA.

WEISS, C. and BUCUVALAS, M. (1980), *Social Science Research and Decision Making*, Columbia University Press, New York.

YUDOFF, M.G. (1979), "Procedural Fairness and Substantive Justice: Due Process, Bureaucracy and the Public Schools", in Newitt, J. (ed.), *Future Trends in Educational Policy*, Lexington Books, MA.

Chapter 17

Indicators, Reporting and Rationality: Understanding the Phenomena

by

Alan Ruby
Department of Employment, Education and Training, Australia

In the hearts of policy-makers and statisticians there is a belief that information changes things for the better. It is a belief that underpins the social sciences, Platonic conceptions of government and models of rational decision-making. The improvement comes because information enhances the capacity to control, monitor and evaluate and, as a consequence, make better decisions and produce better outcomes. This belief shapes all public institutions and most, if not all, organisations and is a dominant principle in the management of resources. Similarly, it has shaped the management practices of most public and private enterprises through doctrines as various as scientific management and "muddling through".

In economics, and particularly in theories of behaviour in the market place, this belief has been most powerful, underpinning the development of national accounts and the creation of an array of economic indicators. As a result, information covering many dimensions of the economy is widely available to allow individuals to compare, evaluate and choose between different products or courses of action.

Information and decision-making in education

In school education the outcome has been different. There has been relatively little information about the outcomes of schools and the ways in which these outcomes have been achieved. There are many explanations for the lack of information on school performance. They range from arguments concerning teachers' professionalism to the apparent difficulties of quantifying the outcomes of education, and they have been buttressed by other dominant beliefs about the nature and organisation of public school systems. In many countries, such as France, Australia, Sweden and the United States, it was assumed historically that the public schools were, for all intents and purposes, uniform; they served populations with similar needs which derived the same benefits from common educational processes. These assumptions supported, and were supported by, a concern to provide common inputs – best illustrated by extensive curriculum regulation, as in Sweden, and the concern for fiscal equalisation in the United States.

In this context, the main data demanded by policy-makers concerned inputs and processes. The use of tax revenues to establish and finance public school systems also supported the demand for data on enrolments and completion rates so as to justify the expenditure on grounds of fairness and efficiency.

Strong policy and ideological commitments to notions of equality and to the social values embodied in the common school serving a local community – such as inclusion and social and racial heterogeneity – interacted with these dominant beliefs and organisational principles. These interactions produced a managerial ethos for public school systems which required relatively little information on performance. Nor was there a need for much information about processes because these were, largely, regulated to ensure uniformity. Monitoring of processes and performance was equated to ensuring that they conformed to regulations and was carried out by political agents or inspectorates which could also provide qualitative information. In this context, there was no need for quantitative information on school outcomes or processes of learning. Instead, the focus was on quantifying resources to ensure common inputs, and measuring stocks and flows of students. There was no need to inform parents about the relative performance of schools or to provide for choice of schools because they were essentially all the same and choice produced inequities (see Lott, 1987, for a discussion on school-zoning and choice).

This notion of uniformity was confirmed by the school and social context studies of the 1960s and 1970s which were popularly interpreted as showing that social background and class variables, rather than schools, determined student achievement (see, for example, Coleman, 1966; Jencks *et al.*, 1972). These interpretations were, in a sense, a celebration of the sameness of schools and schooling.

This perception held sway until the 1980s when it was challenged by studies on school effectiveness which showed that schools could make a difference and that there were differential outcomes, and hence inequities, as a result of school practices and policies. It was also challenged by changes in political philosophy, most notably the extension of the market philosophy and the principle of choice to public services such as education. Demands for comparative information about national school performance in order to address issues of international economic competitiveness and national efficiency have also challenged the orthodoxy. Other broader policy changes have also been influential, notably freedom of information, the devolution of authority and decision-making, and enhanced public accountability. There was a strong connection between these factors and the development of school performance indicators during the 1980s (Ruby, 1989), and they influenced the role accorded to evaluation in national education systems (Granheim, 1990).

A parallel explanation of the changing priority given to data collection and information in education lies in the changing conceptions of educational management. Levin *et al.* (1990) argue that historically the dominant influence on educational management was "scientific management". Its first principle, as enunciated by F.W. Taylor in 1911, was that decision-making should not be intuitive but scientific. It should be based on observation, analysis and inference and guided by the goals of efficiency and rationality. These notions and Weber's later arguments about rational and legal authority and bureaucratic decision-making, designed to ensure equal treatment, supported the development of rudimentary data bases in education.

The political models of decision-making were a counter to these rational influences. Decision-making in these models was to be accomplished by mediating between the various interests rather than by following some logically derived best course of action. In the United States, this model found expression in the local school boards and the strong tradition of elected officials as managers. How relevant it is to other systems is difficult to judge, but the power of interest groups does seem to have limited information gathering in education systems.

Levin *et al.* (1990) write, moreover, that in the 1980s the dominant managerial model has been "applied science": available information is used to identify and define problems and then select a course of action. This model, which subsumes the "reflective practitioner", uses scientific information as well as common sense and practical knowledge. It links information and the quality of education and extends the "demand for the development of more extensive, and more imaginative, educational information systems than planners have used in the past" (p. 78).

While this model emphasizes linking information and quality issues, there has been relatively little scrutiny of proposals and strategies to improve the quality of information about schools and school performance. There are quite a number of major reform programmes in OECD Member countries that aim, variously, to provide better information about schools so as to improve management, governance, accountability and parental choice. They deserve close examination because they represent a significant and substantial break with the prevailing doctrine about information and data in education.

An understanding of what drives and shapes these reforms, many of which are packaged as better "reporting", is needed. What are the key characteristics and underlying rationales that distinguish what is being offered to parents and policy-makers as better reporting? The critical dimensions in such an analysis are: What and whom do these reports cover? Who gets the information? To what extent do they compare students and schools? Do they encompass more complex assessments of quality or are they confined to improving the coverage and presentation of student achievement data? Do they acknowledge the effects of contextual factors and the importance of input and process measures as well as measures of outcomes? Do they include a consideration of net change or value added, *i.e.* the difference in outcomes that can be attributed to schooling?

These questions warrant examination because they bear directly on ideas of accountability and illuminate wider policy questions, such as who controls information about the functioning of schools and where the responsibility for poor performance is located. They also shed some light on the policy context and values associated with the notion of indicators and reporting. Finally, such an analysis makes it possible to take a broader view of these phenomena and the theoretical assumptions that underlie them.

The demand for information

The simplest and most powerful motivation for developing new methods of reporting is the desire to know more about what is happening in the schools. More is involved than curiosity. Mitchell (1989) identifies four main motivations for measuring educational phenomena: *a)* the pragmatic – does it work?; *b)* the moral – is it good?; *c)* the conservative – is it necessary?; and *d)* the rational – does it make things better? All four motivations have shaped educational policy debates in the 1980s, the force of each varying with time, place and policy.

Despite this variance and the number of factors involved, the reality in the 1980s was that many national statistical agencies could not answer the kinds of questions policy-makers and the public were asking about "educational results" (Stern, 1986). There was a lack of coherent and consistent data about what was happening in schools and about the outcomes of schooling. The quality of the data varied greatly and was often poor, with substantial inaccuracies and inconsistencies. The data often had relatively little policy value and only partially covered key policy areas. Information that was available – mostly about financial inputs and enrolments – was of the wrong kind. This was understandable, as the dominant concern of many countries, particularly in the post-war reconstruction period, was the "throughput" of education and the expansion of the system, as exemplified in the French slogan "a classroom a day". In the United States, concern about performance was overwhelmed by attention to graduation and enrolment rates and costs per students and arguments about the efficient use of resources (Murnane, 1987). Outcome data and information about variables linked to outcomes were often lacking (Plisko *et al.*, 1986; Selden, 1986).

The unanswered questions were often those motivated by concerns about the scale of public expenditure on education relative to its effectiveness, the need to reassess public expenditure priorities and to pursue particular policies or reforms. This problem was common to many OECD countries. In Australia, for example, the Quality of Education Review Committee (1985) found "no incontrovertible evidence ... that cognitive outcomes of students were either better or worse" after 15 years of significant federal government funding for school education (p. 49). Existing national data were "largely concerned with mapping of inputs" and could not identify qualitative improvements associated with increased expenditure.

The evidence of outcomes was at the wrong level or had analytical constraints. The large routine assessment of student achievement carried out under that National Assessment of Educational Progress (NAEP) in the United States and the work of the Assessment of Performance Unit (APU) in the United Kingdom were perceived to be inadequate. NAEP's main shortcomings were lack of state by state comparability (Alexander, 1987), the absence of contextual data, and an inability to address particular policy issues. The APU's programme was seen to lack policy relevance and to be pursuing technical questions rather than providing helpful comparative information (Goldstein and Gipps, 1983).

Outside these large-scale assessment programmes, considerable effort had been expended in measuring the effects of specific purpose or categorical programmes designed to address disadvantages or bring about changes in teacher behaviour or school practice through marginal increases in resources. These reviews did not assess schooling as a whole but either assessed progress towards programme objectives which were necessarily limited in scope or tried to ascribe progress towards broader objectives to relatively small resource differentials.

The policy-makers' response to these shortcomings is illustrated by the United States' Secretary of Education's "Wall Chart" and the Australian Federal Employment, Education and Training Minister's advocacy of national reporting "on how well our schools are performing against established goals ... There is currently no single document which informs the citizens of Australia" about the performance of the nation's education systems. Such a report "would raise public awareness of and confidence in schooling and inform policy-makers" (Dawkins, 1988, p. 5). This situation had a parallel in New Zealand. A major review found that lack of information, specifically about students' standards of performance, meant that public discussion was "at best uninformed and at worst destructive for consumer and professional alike." (Picot, 1988, p. 26)

From a different perspective, Finn (1989), commenting on the United States, argued that better outcome data were necessary to challenge the "widespread denial" by American students and their parents of the overall poor standard of achievement. Contrary to general public perceptions, parents consistently report that the schools educating their children are effective and that they are satisfied with their child's performance. Students rank themselves as "good" at mathematics even though this is not borne out by national or international assessments.

The demand for more "relevant" data was reinforced by the availability of highly aggregated data on performance. Murnane (1987) argues that data about economic and educational inequities generated a demand for more information at finer levels of analysis to address strategic and policy questions. A great deal of critical attention was also focused on the weaknesses of the data.

The themes of a lack of information relevant to policy, a population uninformed about school performance, and a skew to input data away from outcome information are repeated in a number of countries (National Center for Education Statistics, 1986). Broader factors such as a concern for economic competitiveness combined with these themes to produce a renewed interest in data gathering, comparative studies of achievement and the development of indicators. This demand for relative performance data is another dominant reason for obtaining "better" information.

While these factors are persuasive at the "macro" policy level, equally strong reasons for obtaining better information exist within education systems themselves. At the school site level, the most common justification for a form of annual public report is accountability and better communication between school and parent about the activities of the school in general or about their child in particular.

At the school level these justifications are often linked to the idea of education as a consumer good that is subject to consumer preference. The argument runs that in order to exercise choice on a rational basis, the consumer must be better informed about the relative quality of schools. The easiest, but not necessarily the best, way to do this is by publishing comparable test results. Versions of this argument can be heard in the United Kingdom, France, New Zealand and the United States, and increasingly also in Australian states. It is most explicit in the United Kingdom, where the government's ideological commitment to the principle of parental choice to bring about "a strong element of competition between schools" (Baker, 1989*a*) makes information about the comparative performance of schools necessary. Without this information "parental choice remains more theoretical than real" (Baker, 1989*b*). In the United States, this argument has been reworked by Chubb and Moe (1990) who argue that choice can "transform" the public education system. They advocate "a true choice system" in which the state assures minimal certification of teachers and schools, parents are able to choose schools freely, and schools are able to choose students and teachers. The choice of school would be aided by parent information centres which would "collect comprehensive information on each school". The "parents and students who directly experience their services and are free to choose" would maintain accountability for "quality of school performance" (Chubb and Moe, 1990, pp. 215-225).

Less powerful in the current climate, but still persuasive, are concerns with equity. Policy-makers, teachers and teacher organisations advocate better reporting about the outcomes for particular groups, such as girls, minority ethnic groups and children living in poverty, so they can pursue broader policy goals and justify differential resource allocation (Ruby *et al.*, 1989). Indeed, better reporting and more comprehensive analyses of achievement by race and gender can inform and support programmes to advance equity goals as well as institutional performance (Lange, 1988).

In addition to these political and ideological justifications, there are professional reasons for improved reporting. Better information about student outcomes and about the processes of schooling can improve teaching practices. At one level, simply knowing how students are performing relative to others in the year or age cohort in the school or nationally can stimulate changes in programmes. Information about resource allocation and use within school programmes can also contribute to school improvement and effectiveness (Ruby and Wyatt, 1989). This of course assumes that communities, teachers and decision-makers are responsive to information and that they have the capacity and the authority to respond.

These professional justifications are emphasized in the new national assessment system in the United Kingdom, which is seen to serve at least three professional purposes: *a)* giving "formative" information which "teachers can use in deciding how a pupil's learning should be taken forward"; *b)* providing "overall evidence of the achievements of a pupil"; and *c)* encouraging professional development by providing a "valuable basis for teachers to evaluate their own work" (Department of Education and Science, 1989). Another rationale offered to justify assessment initiatives is that they provide teachers with diagnostic information and information that will assist them to "recognise where improvement is needed" (NSW Department of Education, 1989*a*, 1989*b*).

Better information about performance and the processes of schooling is also expected to benefit students themselves. Most such benefits are the product of assessment procedures rather than direct results of broadly-based reporting strategies. Under the national assessment system run in the United Kingdom, for instance, the results will give students "clear and understandable targets and feedback about their achievements" (Department of Education and Science, 1989).

The demand for better information has broader motivations than simple pragmatism and morality. Policy-makers press for data about outcomes so as to address questions of efficiency, competitiveness; they need information to support wider reforms, notably in the areas of greater accountability and choice. There is, however, a rational element to these factors; it is the assumption that improvement in outcomes and professional practice will follow from the provision of information and the application of reforms and policies.

How these forces shape practice varies according to the traditions and policy imperatives of individual countries. These variations provide a means of gauging the true direction and nature of these forces. Similarly, studying application rather than argument provides a better measure of the strength of a particular factor relative to other values or policies governing or underpinning national education arrangements.

Some country examples

In the United States, in addition to supporting statistical and indicator work on the conditions of education and the well-being of youth, the most obvious impact is in student assessment, particularly in the coverage and use of outcomes data. The new reporting "ideology" challenges the practice of the light sample, with no comparisons and no disclosure of information to the general public. The most notable changes are in the NAEP which has adopted state by state comparisons after 20 years of restrictions. To make these comparisons, assessments have to include more schools and more students. The NAEP Review Panel also favoured a closer link between the tests and core curriculum content and an assessment of a range of outcomes including acquisition of "higher order" skills (Alexander, 1987, p. 8).

Student assessment has also been the focus in the United Kingdom, where the Education Reform Act of 1989 introduces national assessment at ages 7, 11, 14 and 16 years, with some limitations on testing at the end of the first stage. Increases in scale are matched by increases in scope and in material tested. The national strategy emphasizes the importance of assessing a wide range of outcomes by having attainment targets across all aspects of seven foundation subjects for different levels of schooling and ability. There is a similar breadth of coverage in the New South Wales (Australia) basic skills tests for Year 6 which assess aspects of literacy and numeracy – reading and language skills, numbers, measurement and space.

In addition to widening the scope of assessment, there have been changes in the use of the results, notably the disclosure of test results to parents. In New South Wales, individual reports are issued to the parents of every child tested to provide "information about the attainments of their own children in the basic skills". There are similar provisions in the United Kingdom, where the examination results of each school have been made public since 1980. Comparison has been difficult, partly because of the way the figures are presented and partly because of the number of separate examining boards assessing achievement at the end of secondary schooling. The reduction of the number of examining boards since 1988 and new regulations requiring, from 1991, a more straightforward presentation of figures will make this information more accessible. Since the 1980 policy change, some local education authorities have developed sophisticated methods of harmonizing the different results and reporting to schools and the wider community. Recently, school by school results on the *baccalauréat* examination have been published in France. The data are highly aggregated and arranged alphabetically by suburb and town, making comparison difficult.

In Denmark, the same policy concerns have produced different responses. The government's goal of promoting choice and competition in the school sector has raised questions about how schools should be evaluated. There is no publicly available outcome data to compare schools, and high school examination results are not published. In response, the Ministry has introduced two pilot projects in school evaluation: the quality school strategy and the self-evaluation programme. The latter encourages schools to systematically evaluate their own operations while the quality school model begins with a self-evaluation followed, six months later, by a week-long inspection visit that includes discussions with teachers, students, parents and the community. A draft report is produced and presented at a public meeting. Amendments of fact but not judgement are negotiated, and the report is usually published by the school. While fundamentally different from the reporting based on student assessment regimes, this model – based on

BEST COPY AVAILABLE

the classic notion of expert judgement – underscores the growing commitment to public release of data for comparing and judging schools.

The Norwegian experience is similar. With a highly decentralised and devolved school system and a strong commitment to equality of access and standards, Norway has not had a tradition of "accountability" and evaluation; indeed, the word accountability has no direct equivalent in Norwegian. The situation has begun to change, due to debate about the link between the national economy and educational standards and to the pressure for all public institutions to justify how they use public monies. The National Council for Upper Secondary Education has responded by developing a pilot project on school evaluation which can "contribute to the healthy development of individual schools while at the same time meeting the needs of authorities for information about the work of schools".

This project focuses on involving parents, teachers and educators in a "systematic and long term analysis" which aims to "assess the results of the educational and subject oriented process, the social environment as well as relations between the school and the community". One of the key features of the pilot project is the production of annual reports which schools prepare according to a general outline. These reports have created some apprehension in schools due to fears that the central authorities will use them to justify intervention in schools. Another substantial concern is that this evaluation approach may produce a national evaluation system too slowly and that an "inappropriate system of testing" may be imposed (Björndal, 1990, p. 16).

Responding to similar pressures and to the findings of an OECD review, the Norwegian Ministry of Education instituted project EMIL to develop "a system of steering by goals and evaluation". The key findings were that central planning documents should be more concise and present clearer and more concrete goals to allow for greater flexibility at the local level, to define responsibilities, and to provide a basis for evaluating the results. The evaluation will have four functions: "to control the quality of education, to provide a basis for decisions and changes, to make the education visible [and] to provide motivation." (Granheim and Lundgren, 1990, p. 40)

In their approaches to evaluation and reporting, the Norwegians and the Danes seem to give greater weight to local responsibility and consequently to place greater trust in local and school authorities to make improvements. They continue to be unwilling to compare schools and to be doubtful about the benefits of national measurement of student outcomes. They do, however, share the belief that evaluation and the public availability of more information about schools will stimulate and provide a basis for improvements in quality. What is unclear in this model is how more information can or will be provided, and by whom, without conflicting with the underlying traditions of school self-evaluation and local and regional autonomy. This tension is inherent in the outcomes of the EMIL project which leave unasked the question: "Which social values should a national agency monitor through evaluation or statistical collection?". In relatively small countries with homogeneous populations and a high degree of social cohesion, the ambiguity is accommodated by loose coupling of the parts of the system and, in the case of Norway, diffusion of responsibility for evaluation across many relatively small agencies.

In contrast with their Nordic neighbours, successive Swedish governments traditionally promoted a strong central model of control and regulation. In the last ten years, significant reforms have devolved authority for education to regional and local levels.

Presiding over these reforms has been the careful development of a national evaluation strategy.

The cornerstone of this strategy is a programme of student assessment. Swedish educational researchers have a long and highly respected involvement in international studies of student achievement (Husén, 1990). National assessment was introduced after ten years of investigation, discussion and research drawing on this international experience. The key principles were that the assessment should involve knowledge and skill tests that encompass a total curriculum, study process and yield, cover all compulsory schooling, and be formative by offering guidance to teachers by assessing mid-points rather than end points in stages of schooling.

The first assessments in 1989 were made in grade 2 (Swedish, mathematics, music and art) and grade 5 (Swedish, mathematics, music, art, civics, science and English). In 1992, student assessments will be made at a third point, which was to be grade 8 but will now be grade 9, the final year of compulsory schooling.

The results of the 1989 tests were made available to teachers who received their own results individually and in comparison with the national results. School results on these tests have been published but have received little publicity.

Triennial school reports to and annual qualitative studies by the 24 regional boards complement this form of assessment. The qualitative studies look at issues such as how time is used. Another complementary strategy is age cohort studies, with five separate tracks. All of these evaluative devices will come together in a triennial report to the Government on the conditions of schooling. The Government is to respond a year later setting out its programme of action and priorities.

This is a very rich programme of evaluation and monitoring; it looks at processes and outcomes and balances qualitative and quantitative data. It includes clear lines of responsibility and authority, thus allowing considerable freedom for local variations and differences while retaining a means of assessing progress towards national goals. More recent reforms that further reduce the central administration and its responsibilities may change the nature of reporting and control arrangements in Sweden, but clearly defined responsibilities for the various levels of government seem likely to be maintained.

To date, the Nordic countries have been less influenced by public expenditure constraints and the need to answer questions such as: "How much are we spending?", "Why is it that some students are failing if we are spending so much?" and "Why are there no appreciable increases in student performance over the years although there have been real increases in financial inputs?". These questions concern policy-makers in the United States and other countries and underscore the importance of placing outcome data in a context that promotes informed debate.

None of the new reporting approaches are designed to answer these questions by measuring the "value added" to the student, that is the net gain as a result of schooling rather than simply the final outcome. The notion of value added is based on the premise that because students come to school with different experiences and levels of knowledge and because some schools have higher concentrations of students less prepared for schooling, these factors should be taken into account when comparing student achievement across schools. Value-added comparisons give a truer picture of the effectiveness of the processes of schooling. Factors such as race, language background, and special needs can also be applied in the interpretation of student outcomes. While there are considerable technical challenges involved in adjusting student assessment scores for these different background variables, there have been some successes in recent years with sophisticated multi-level modelling approaches (see Nuttall *et al.*, 1989; Gray *et al.*, 1990). In addition, considerable interpretative value can be obtained from descriptive information – about neighbourhoods and educational policies and practices, for example – that can be manipulated to improve outcomes and available inputs and to place outcome data in a broader context.

This contextual information is present in some of the reporting schemes: they also collect data on other aspects of schooling – retention rates, student absenteeism, teacher qualifications, participation rates in various subjects for boys and girls, and student or home background characteristics, such as ethnicity and time spent watching television and doing homework. Financial data on school expenditure and costs of various categories of provision and related measurements, such as class size and student:teacher ratio, are also collected. This information is often already available to policy-makers but not always systematically or routinely. It is seldom accessible to the public. Routine systematic reporting of this information to the public, to the school and the system has the potential to increase public understanding of the strengths and weaknesses of the educational enterprise as a whole. The underlying assumptions are that parents really do want to know what is happening and will use their knowledge either to call the school to account or to determine where their child will be educated.

What happens to the information: who uses it?

Most of the debate about the use of information in education has concerned public disclosure of test results. Issues raised include the impact on curriculum, "teaching to the test", and overemphasis on lower-order skills; some have been studied by Nuttall (1988). While they are important to the work of schools and teachers because they concentrate on the emotional issues of testing and standards, these issues shift attention away from other, more significant concerns: the relationship between parent, school and teacher and the responsibilities of decision-makers to respond to evidence of school failure.

While there are significant differences among countries and cultures, most school systems tend to circumscribe the role and rights of parents in relation to schooling. The very notion of compulsory schooling, for example, places the parent in a subordinate relationship to the school. Similarly, ideas about professional autonomy, the teacher standing *in loco parentis* and the teacher's accountability to the wider community through hierarchical structures such as local authorities and State agencies have, indirectly, distanced parents from schools and teachers.

Accessible information about the goals and outcomes of schooling, and particularly about the relative performance of individuals or schools, starts to change this relationship. It gives parents a basis on which to begin a dialogue with schools and teachers about student progress and how it might be maintained or improved. The act of giving this information to parents implies that the teacher or school is or should be accountable to parents. This shift in relationships is reinforced by publicly endorsed statements that parents have the "right" to know how their child is performing and by the emphasis on the parent as a "consumer" who can exercise choice in school services.

Availability of information on such factors as staff qualifications, professional development, budget and curriculum organisation also influences the relationship between parent and school. It broadens the scope of the accountability relationship to encompass the major elements of a working model of the school education process. Parents can see how fees or taxes are being spent, the range of provisions which are necessary to the educational process, and the policy and organisational framework which is shaping their child's education.

The powerful element in the shift in the relationship is that parents are thus better equipped to make judgements about the schools and teachers that serve them. This would have very great impact on schools and how they are organised. It may mean that schools and teachers abandon many of their current, highly valued practices for assessing students' work and communicating with parents and replace them with formal and structured reporting procedures. This could depersonalise relationships between teachers and parents. Teachers may also temper their qualitative assessments either to limit the accountability relationship or to restrict the capacity for comparison. Both responses are understandable.

Alternatively, information of sufficient breadth may create greater appreciation and support from parents and the public. The extent of the impact on the work of schools and teachers will depend largely on if and how parents take up and use this information.

At the level of overall educational policy-making similar questions arise. What policy responses can and should be made to "poor performance"? The evidence available shows that this is not an easy task.

In Sweden, evaluation strategies are purposely designed to reinforce structural changes in the governance of education. Recent decisions to further decentralise the management of schooling are based on the assumption that parents and local politicians have the capacity to make the necessary decisions. To exercise this capacity they need information, about their schools and about the national standard, and the necessary skills and motivation. The former Swedish Minister believes that parents are "able to demand their rights in a way that was not possible previously" because earlier reforms have produced an education "which not only teaches what democracy is, but ... also teaches the most important lesson, how to use democracy" (Persson, 1990). The benefits to be gained will provide the motivation.

To implement these new reporting strategies, decision-makers at the system, regional or provincial level will require new skills and have to adopt different managerial practices. They will especially need to be more open about what they do and share information with the wider community. They too must become more accountable – more precise about the purposes and costs of activities. Most of all, they will have to accept being judged relative to past performance and in comparison with other systems.

A crucial aspect of the accountability relationship at these levels is decision-makers' responsibility to assist a school that is "failing". In the United States, some elements of the education reform movement of the 1980s advanced the idea of the educationally bankrupt school or district (Committee for Economic Development, 1985, p. 29), and this has been adopted in some states and districts. In New Zealand, the new Review Agency has the capacity to dismiss the Principal and/or the Board of Governors of schools that are not fulfilling their charter. These are punitive responses which will not necessarily create improvement. They are also essentially organisational responses. They are not policy responses based on improved information.

It is an open question whether and how policy-makers will synthesise indicator information in decision-making processes. Nonetheless, the new reporting strategies potentially have a substantial capacity to inform policy-making. The use of goal-based measures, the strong ties between key policy values and the variables being monitored, and the emphasis on contextual and outcomes measures all contribute to making the products of these strategies relevant to policy. How much they will contribute to policy-making will not depend solely on the utility or validity of the information. Indeed, there is no certainty that the information will influence policy-makers at all.

The Swedish model addresses these questions quite directly by providing both for a national government response to the "conditions of schooling" and by placing greater responsibility with the school and local authorities. The first national assessment survey offered findings that were useful in policy-making. For example, the survey showed that the two main methods of teaching reading produced no differences in outcomes, and this allowed the national authorities to accept both and hence neutralise a divisive and unproductive debate.

In developing countries, improved information has led to some changes in resource allocation and decision-making in the system. Recent examples are documented in Ross and Mahlck (1990). Most are relatively direct responses to statistics describing specific conditions such as physical facilities, or to observations of teaching practice. Similarly, there have been conventional policy responses to simple analyses of test scores, as in the case of SAT scores in the United States. However, "back to the basics", increasing curriculum time in science and mathematics, and spending more money have not been shown to increase student outcomes (Steelman and Powell, 1985; Finn, 1989). Other commentators point out that changes in measured outcomes may be due to the successful pursuit of other equally or more highly valued policy goals, particularly social goals. Howe (1985), for example, argues that the decline in SAT scores from the 1960s to late 1970s was due, in part at least, to changes in the composition of the population taking the test due to efforts to improve access for Blacks and Hispanics to educational services. For reasons such as this, national assessment information has not always contributed to policy-making (Dockrell, 1988, p. 16).

This history of mixed responses to the provision of information underscores the fact that even where information is timely, accessible and pertinent, it is still only one element in the decision-making process.

A celebration of rationality?

At first glance education indicators seem to be a triumph of rationality. It is possible to discern in the various reporting schemes – both those discussed here and in Wyatt (1990) – the basic tenets of scientific management and rational authority. The same is true for the various models of indicators surveyed by van Herpen (1992). It is also possible to see in the reporting proposals a commitment to notions of competition and regulation through the market: data will inform choice and there will be closer and more authoritative relationships between institutions, such as schools, and clients or parents. These ideas draw their power, and a measure of legitimacy, from economic theory, notably from the assumption that individuals will act rationally to maximise benefits when they make choices and exercise preferences in a range of circumstances. The existence of perfectly correct information about the current situation and likely outcomes allows the individual to "act rationally" to select the behaviour or set of circumstances most "appropriate or apt relative to (his or her) presumed wants or needs." (Bennett, 1964, p. 84)

The initiatives described above also rely on information. Indeed, the demand for it is very strong because information is necessary to the fundamental reform: the shift away from regulated systems and regulated teacher behaviour. Information is necessary to select courses of action primarily on the basis of reasoning, professional judgement or "good government" rather than one management based on rules. These initiatives are thus based on a belief that people will act reasonably if they have enough information on which to base their decisions. This belief underlies proposals for parental choice of school, the structural changes proposed in Sweden, and the school improvement programmes implemented in many OECD Member countries.

It is a strong belief that carries with it a concept of the individual, a model of decision-making, and a theory of democracy. For these reasons, it is worth fostering and pursuing. However, it also carries dangers. The most obvious is the risk of being caught in a formalist construct – simplistic Taylorist scientific management and authority accorded to quantitative information. This leads to denying the importance of qualitative information and of the tasks of choosing, defining and agreeing on the goals and values of education, which are essential parts of public policy-making, the management of schools and the conduct of learning. It also results in ignoring the role of other forms of knowledge in problem-solving – casual and common knowledge, non-scientific and non-professional ways of knowing and analysing – common sense, casual empiricism and thoughtful speculation (Lindblom and Cohen, 1979, p. 12).

Equally important, but at another level, overemphasis on information and rational analysis disregards the possibility, or the probability, that problems can be solved by other routes. It ascribes to rationally derived information the status of "proof" or final judgement rather than of evidence to be tried, tested, discounted or accepted in the search for a possible or desirable course of action.

Even when these dangers are avoided, there are problems to be resolved if very rationalist models of information systems and decision-making are to be applied to schools. These problems are embedded in the theoretical assumptions which characterise rationalist models, particularly those based on economic positivism. These assumptions include the belief that individuals will always pursue their self-interest and that they are, with some allowance for indifference, able to perceive the best way to realise those interests. Underlying these assumptions is the related idea that all action is purposive, and that changes in preferences or inconsistency of choice can be ascribed to changes in information or expectations. In short, the chain of assumptions is that the informed individual, knowing what to achieve, chooses the best or optimal combination of goods or services (see Caldwell, 1982, p. 139 ff.).

In traditional economic analysis, these choices are often described by utility and preference functions. These functions show the relationship between competing variables, such as income and leisure or school fees and teacher qualifications. The trade-off between these variables is described as the "indifference curve", the series of points where the utility to the individual is the same. The point of "balance" between the variables depends on the individual's values and preferences, but the utility and preference functions predict the likely range of combinations.

Dispassionate analysts will find it difficult to apply this reasoning to parents exercising their preferences in situations such as schooling, where there are:

- multiple decision-makers who affect the choice of school and school activities;
- multiple and contested definitions of what is best or good;

- a high degree of tolerance about how to fulfil the various definitions of what is best or good;
- less than perfect and incomplete information about the commodity or service; and
- overt homogeneity despite the apparent capacity for variation in the services provided by schools.

In short, public school systems are not readily perceived as an economic good for which choice can be adequately predicted by means of preference or utility functions. If this is so, then there are limits on the power of information and hence on choice as tools for reform and improvement.

Such limits do not, by themselves, invalidate the demand for information or question its legitimacy. Indeed, recognising these limits may make it possible to define more precisely the functions and explanatory force of information. It will also help analysts and decision-makers recognise that the power of information is partly determined by broader social relationships: models of decision-making, models of government and the balance of authority within schools, and between schools, students and parents.

Conclusion

The above discussion underscores the need for methodological pluralism. Caldwell's arguments about economics are equally valid in this context: we must recognise that there is no "single, universal, prescriptive scientific methodology" (1982, p. 244). In education, the simple models and theoretical frameworks guiding work on indicator development offer a promise of making progress. Even if they fail, they do not disprove the basic proposition: better information makes for better decisions.

Bibliography

ALEXANDER, L. (1987), "The Nation's Report Card: Improving the Assessment of Student Achievement", *Report of the Study Group*, National Academy of Education, Boston.

BAKER, K. (1989*a*), Address by the British Secretary of State for Education and Science, *Conference on Secondary Education*, Sydney, 10 May.

BAKER, K. (1989*b*), "Education Reform – An International Issue", Press release by the British Secretary of State for Education and Science, Sydney, May 10.

BENNETT, J. (1964), *Rationality: An Essay Towards Analysis*, Routledge, London.

BJÖRNDAL, I. (1990), "National Aims – Local Freedom: A School Evaluation Project in Norway", The National Council for Upper Secondary Education, Oslo.

CALDWELL, B. (1982), *Beyond Positivism: Economic Methodology in the Twentieth Century*, Allen and Unwin, London.

CHUBB, J. and MOE, T. (1990), *Politics, Markets and American Schools*, The Brookings Institution, Washington, D.C.

COLEMAN, J.S. (1966), *Equality of Educational Opportunity*, U.S. Government Printing Office, Washington, D.C.

COMMITTEE FOR ECONOMIC DEVELOPMENT (1985), *Investing in our Children*, Research and Policy Committee, Washington, D.C.

DAWKINS, J.S. (1988), *Strengthening Australia's Schools: A Consideration of the Focus and Content of Schooling*, Australian Government Printing Service, Canberra.

DEPARTMENT OF EDUCATION AND SCIENCE (DES) (1989), *National Curriculum: From Policy to Practice*, DES, United Kingdom.

DOCKRELL, W.B. (1988), "National Quality Indicators", *Scotland Studies in Educational Evaluation*, Vol. 14 (1).

FINN, C.E. (1989), "A Nation Still At Risk", *Commentary*, May.

GOLDSTEIN, H. and GIPPS, C. (1983), *Monitoring Children: An Evaluation of the Assessment of Performance Unit*, Heinemann, London.

GRANHEIM, M. (1990), "The Education System in Norway: The Need for Evaluation", in Granheim, M., Kogan, M. and Lundgren, U.P. (eds.), *Evaluation as Policy-Making*, Norvegian University Press, London.

GRANHEIM, M. and LUNDGREN, U.P. (1990), *Steering by Goals and Evaluation in Norwegian Education*, NORAS, Oslo.

GRAY, J., JESSON, D. and SIME, N. (1990), "Estimating Difference in the Examination Performances of Secondary Schools in Six LEAs: A multi-level approach to school effectiveness", *Oxford Review of Education*, Vol. 16 (2).

HOWE, H.W. (1985), "Let's Have Another SAT Score Decline", *Phi Delta Kappan*, May.

HUSÉN, T. (1990), "Swedish Education in an International Perspective", in Harnqvist, K. and Svensson, N. (eds.), *Swedish Research in a Changing Society*, Gidlunds, Varnamo.

JENCKS, C. *et al.* (1972), *Inequality: A Reassessment of the Effect of Family and Schooling in America*, Basic Books, New York.

LANGE, D. (1988), *Tomorrow's Schools: The Reform of Education Administration in New Zealand*, Wellington, August.

LEVIN, H., WINDHAM, D. and BATHORY, Z. (1990), "The Relationship between Educational Information, Educational Management and the Quality of Education", in Ross, K.N. and Mahlck, L. (eds.), *Planning the Quality of Education: The Collection and Use of Data for Informed Decision-making*, UNESCO and Pergamon Press, Oxford.

LINDBLOM, C.E. and COHEN, D.K. (1979), *Usable Knowledge: Social Science and Social Problem Solving*, Yale University Press, New Haven.

LOTT, J. (1987), "The Institutional Arrangement of Public Education. The Puzzle of Exclusive Territories", *Public Choice*, Vol. 54, pp. 89-96.

MITCHELL, D.E. (1989), "Measuring Up: Standards for Evaluating School Reform", in Sergiovanni, T.J. and Moore, J.H. (eds.), *Schooling for Tomorrow: Directing Reform to Issues that Count*, Allyn and Bacon, Boston.

MURNANE, R.J. (1987), "Improving Education Indicators and Economic Indicators: The Same Problem?", *Educational Evaluation and Policy Analysis*, Vol. 9 (2).

NATIONAL CENTER FOR EDUCATION STATISTICS (NCES) (1986), *The Condition of Education: A Statistical Report*, Office of Educational Research and Improvement, U.S. Department of Education.

NEW SOUTH WALES DEPARTMENT OF EDUCATION (1989*a*), "Basic Skills Testing Program in Years 3 and 6 – Notification of the Date of Testing", Memorandum to Primary Principals, Director-General of Education, 9 March.

NEW SOUTH WALES DEPARTMENT OF EDUCATION (1989*b*), "An Overview of the Basic Skills Testing Program", *Basic Skills Testing Program Information Bulletins*, No. 1, March.

NEW SOUTH WALES DEPARTMENT OF EDUCATION (1989*c*), "Some Common Questions About Assessment", *Basic Skills Testing Program Information Bulletins*, No. 2, May.

NUTTALL, D.L. (1988), "The Applications of National Curriculum Assessments", *Educational Psychology*, Vol. 8 (4).

NUTTALL, D.L., GOLDSTEIN, H., PROSSER, R. and RABASH, J. (1989), "Differential School Effectiveness", *International Journal of Educational Research*, Vol. 13 (7).

PERSSON, G. (1990), "Teacher Policies for the 1990s", Address to a Conference on Teachers and Teaching, Stockholm, September 20-21, Wijk, Sweden, mimeo.

PICOT, B. (1988), "Administering for Excellence", Report of the Task Force to Review Education Administration, Wellington.

PLISKO, V.W., GINSBURG, A. and CHAIKIND, S. (1986), "Assessing National Data on Education", *Educational Evaluation and Policy Analysis*, Vol. 8 (1), pp. 1-16.

QUALITY OF EDUCATION REVIEW COMMITTEE (QERC) (1985), *Quality of Education in Australia – Report of the Review Committee*, Australian Government Printing Service, Canberra.

ROSS, K.N. and MAHLCK, L. (eds.) (1990), *Planning the Quality of Education: The Collection and Use of Data for Informed Decision-making*, UNESCO and Pergamon Press, Oxford.

RUBY, A. (1989), "Why Invest in Performance Indicators?", Bulletin 11 in the Reporting on Educational Progress Monograph Series, Conference of Directors-General of Education, Sydney.

BEST COPY AVAILABLE

RUBY, A., MILLAN, M. and HOWE, V. (1989), "Indicators for Equity: A Neglected Part of the Debate", Bulletin 13 in the Reporting on Educational Progress Monograph Series, Conference of Directors-General of Education, Sydney.

RUBY, A. and WYATT, T.J. (1989), "Why Invest in Performance Indicator Development", in Ruby, A., Wyatt, T.J., Norton, S., Davies, B. and Shrubb, S., *Reporting Educational Progress: Performance Indicators in Education*, Australian Conference of Directors-General of Education, Sydney.

SELDEN, R.W. (1986), *Strategies and Issues in the Development of Comparable Indicators for Measuring Student Achievement*, White Paper, State Education Assessment Center, Council of Chief State School Officers, 30 April.

STEELMAN, L.C. and POWELL, B. (1985), "Appraising the Implications of the SAT for Educational Policy", *Phi Delta Kappan*, May.

STERN, J.D. (1986), *The Education Indicators Project in the US*, Center for Education Statistics, Washington, D.C.

van HERPEN, M. (1992), "Conceptual Models in Use for Education Indicators", in *The OECD International Education Indicators: A Framework for Analysis*, OECD/CERI, Paris, pp. 25-52.

WYATT, T. (1990), "Indicators, Evaluation and Accountability Mechanisms in Public School Systems", Bulletin 15 in the Reporting on Educational Progress Monograph Series, Conference of Directors-General of Education, Sydney.

BEST COPY AVAILABLE

Chapter 18

Policy Uses and Indicators

by

Linda Darling-Hammond
Columbia University, United States

Education indicator movements have waxed and waned for more than a century, as have political concerns over education. Eras of intense interest in indicator development have often been closely linked to governmental monitoring and policy agendas. This chapter examines how indicators have been and may be used to serve policy ends, what the potential benefits and dangers of various uses are, and how guidelines for appropriate relationships between indicators and policy might be forged.

The chapter's perspective on indicator uses and potential abuses is informed primarily by policy approaches in the United States and by the intense political and technical debates that have characterised indicator development there. Although the governance, management, and philosophical foundations of the education system of the United States differ in some important ways from those of other countries, many of the themes treated here have appeared in discussions emanating from governments and education systems around the world.

The relationship of education indicators to policy

Indicators can be simply defined as individual or composite statistics that reflect important features of a system, such as education, health or the economy. Their "overriding purpose ... is to characterise the nature of a system through its components, their interrelations, and their changes over time. This information can then be used to judge progress – towards some goal or standard, against some past benchmark, or by comparison with data from some other institution or country" (Shavelson *et al.*, 1989, p. 4). Thus, indicators are intended to be evaluative in nature, not merely informative: "Statistics qualify as indicators only if they serve as yardsticks." (p. 5)

Of course, anything that is inherently evaluative involves value judgements. Decisions about education indicators are determined with reference to the goals of various actors within society. Since competing conceptions of curriculum goals dictate different topics for investigation (Eisner and Vallance, 1974), changes in a society's views of its educational needs and goals (or in the views of the political party predominating at any given point in time) can, and do, influence the nature of data collected and the ways in which they are used. Because of this, governmentally controlled indicators have tended to reflect each historical era's predominant political and social ideologies, and measures have been developed in response to (or occasionally in reaction against) prevailing goals for schooling.

National goals for education

Of particular importance for the course of indicator development in education is the fact that public schools are created primarily to meet the state's need for an educated citizenry. State goals include: socialisation to a common culture (education to meet social needs); inculcation of basic values; preparation for citizenship (education to meet political needs); preparation for occupational life (education to meet economic needs, including the acquisition of those skills and academic disciplines thought to be necessary foundations for later productive life). Because the public school is embedded in an egalitarian legal system, equality of educational opportunity is an important value that has become – in some societies and in some eras – both a goal of education and a measure of its processes and outcomes (Wise and Darling-Hammond, 1984, pp. 1-2).

Of course, schools concern themselves more broadly with students' cognitive and psycho-social development, and they frequently embrace a more comprehensive view of academic learning. Many parents, educators, and policy-makers value learning for its own sake and as a means to individual growth, aside from its social uses. Nonetheless, the rationales underlying the creation and public funding of schools have to do with the social benefits of education, not its contribution to private, individual growth. Compulsory education requires students to attend schools for society's benefit; because society stands to gain from at least certain kinds and amounts of widespread schooling, education is an obligation exacted of young citizens as much as it is a right or a privilege.

To the extent that indicator systems are born of the need to monitor the degree to which schools meet society's goals, they are more likely to measure those aspects of schooling expected to be instrumentally associated with economic, political, and social goals – such as student performance in mathematics, science, reading and citizenship – than with other areas of cognitive development and personal growth – such as the inculcation of a love of learning, the development of artistic and aesthetic capacities, or the enhancement of individual resourcefulness.

Measuring the accomplishment of societal goals necessarily requires some reduction in curriculum goals. To evaluate the acquisition of that "certain amount of positive knowledge" described by a nineteenth century school superintendent as essential (Tyack, 1974, p. 49), someone must define precisely what this knowledge is. The definitions proposed by many authors in the course of history have often tended to be narrowly utilitarian (Cremin, 1961), focusing on discrete pieces of knowledge and skills or on types of behaviour. Some definitions have excluded kinds of knowledge or action seen as beneficial to the individual but not as directly useful to the society. As one American educator put it in 1909:

> "Ordinarily a love of learning is praiseworthy, but when this delight in the pleasures of learning becomes so intense and so absorbing that it diminishes the desire and the power of earning, it is positively harmful. Education that does not promote the desire

and power to do useful things – that's earning – is not worth the getting." (Callahan, 1962, p. 10)

This point of view, which prevailed in the construction of indicator systems during the turn-of-the-century "scientific management" era, was criticised by John Dewey, among others, as ultimately self-defeating because of its narrowness and shortsightedness (Dewey, 1904). Nonetheless, indicators directed specifically at measuring the prospects for wage-earning success in jobs both reflected and directed the education systems of the times. Whereas tests became the instruments for "tracking" students in the United States, examination systems were constructed in many European and Asian countries to allocate opportunities for further academic study or for vocational training.

Curriculum goals

The thorny issue of curriculum goals has been debated over the last century or more during which education has been observed and shaped by indicators. Governmental needs for information and data also influence the design of indicator systems and the potential uses of indicators for policy decisions. The purposes of indicator systems include at least the following: monitoring the general conditions and contexts of education; identifying progress towards specified goals; illuminating or foreshadowing problems; diagnosing the potential sources of identified problems.

In each case, of course, decisions must be made concerning the features of the system that will be examined and the goals that will be deemed important, along with the perceived problems and their potential sources that will be explored. Earlier in this century, for example, school drop-out rates were not viewed as a major problem, since most of those who left school could still become gainfully employed. Now that this is no longer the case, drop-out has become a policy issue.

Given agreement about a policy problem, there will still be many questions to answer, such as how to monitor, examine, and understand the problem and its sources. Should one examine the characteristics of children likely to drop out and those of their families, or should the research be focused on the aspects of schooling that may produce dropping out? In examining schools, precisely what should one look at? The questions and answers differ, depending on the views of schooling and social goals that are adhered to. Decision-makers and researchers may choose to explore school policies and practices concerning curriculum, instruction, student grouping, placement, and promotion as relevant to a given problem, or they may not. They may choose to monitor non-school factors such as television viewing instead. The variables selected for examination will influence the choice of strategies and the chances of addressing a given policy problem constructively.

The influence of politically derived and used education indicators is felt, because they direct the attention of educators and the public to the specific concerns suggested by published data and to the interpretations of those data inherent in their presentation. Other presentations may suggest different views of both the sources of problems and their solutions. Decisions about what curricular content, processes, and outcomes should be measured can have far-reaching implications for what is taught and learned in schools, as well as what is studied and transmitted in research communities.

The indicators themselves may be used as arbiters of incentive systems, as is the case, for example, in countries with a policy of using test scores for allocating rewards to schools or for conferring or denying educational opportunities to students. These uses lead to actions that are intended to influence the measurements, thus redirecting curriculum goals and/or reshaping other educational processes. Research in several countries, including Japan, Ireland, England and Belgium, has demonstrated the effects of "high stakes" examinations on curriculum and teaching (Haney and Madaus, 1986).

Policy functions of indicators

There are many important policy functions for which indicators might be used: the pursuit of equity goals, the pursuit of accountability, setting priorities or generating options for policy agendas, and the assessment of goal attainment. An intelligently constructed and interpreted indicator system that comprehensively monitors key features of an education system could be helpful for providing information on most if not all of these functions. However, as explained below, using indicators for accountability presents many difficulties and challenges.

It is most important that the construction and interpretation of the indicator system should be based on a model of the education system, and on an understanding of its operations, which are derived independently from any particular policy agenda. Moreover, as Bryk and Hermanson note (Chapter 2), important differences exist in the conceptions that undergird competing models of schooling. These conceptions must be explicitly addressed in the development of an indicator system. Indicators must be stable over time; they must allow the evaluation of many competing hypotheses, as well as the broad monitoring of trends in a variety of areas which may not be viewed as central to any particular policy agenda at a given point in time. As Eide puts it: "In order to judge what is going on, what is needed are indicators which provide insight into the total process." (1987, p. 8)

If indicators were to be viewed as fully interactive with policy needs – such that current policy agendas determine the choice of indicators to be monitored, and indicator "findings" determine future policy agendas – then it could be predicted that their utility for informing educational policy-making would decrease. The vicissitudes of political ideologies and agendas, along with the limitations of national governmental policies as agents of educational improvement, would render an indicator system too narrow and too changeable to be of much diagnostic use in the long term.

These observations are not meant to be an indictment of the political process or of the necessarily complex and important work of policy-making. Politics engages values and negotiates goals. Policy-making and, for that matter, indicator development cannot honour the views of diverse interest groups unless it takes place in a vigorous political process in which values are proffered, tensions brought to the surface, and dilemmas fully engaged. The development of indicators, therefore, cannot be divorced from policy-making. In the long-run service of a democratic political process, they must be constructed so as to accommodate many policy agendas, and to provide information on additional matters not clearly central to any policy agenda.

If policy-makers really want to understand what is going on in an education system, they will need not only a comprehensive set of indicators, but also an extensive research portfolio examining teaching, learning, and policy implementation in schools. As Ruby notes, "What indicators do is to identify areas for close examination and analysis" (1989). Stern and Hall agree: "No one set of indicators will provide a definitive measure of a programme; they will identify areas where action or extensive qualitative research is necessary." (1987, cited in Ruby, 1989)

Although social indicators aim to monitor the health or status of an enterprise, largely in order to inform policy, their history indicates that promises of direct policy applications have been misleading. They cannot provide the kind of detailed information necessary for answering fundamental questions or evaluating specific programmes or policies. Neither can they define what good education or desirable legislation is (Shavelson *et al.*, 1989; de Neufville, 1975). What social indicators may do reasonably well is to:

> "... describe the state of the society and its dynamics and thus improve immensely our ability *to state problems in a productive fashion, obtain clues as to promising lines of endeavour, and ask good questions* ... The fruits of these efforts will be more directly a contribution to the policy-maker's cognition than to his decisions. Decision emerges from a mosaic of inputs, including valuational and political, as well as technical components." (Sheldon and Parke, 1975, p. 698)

With this perspective on indicators as one of many inputs to "reflective policy-making" in mind, one can turn to questions such as: How can indicators be used for improving policy and practice? What are the issues and difficulties involved in using indicators constructively?

The potential of indicators for changing policy and practice

If indicators are seen a guide to reflective policy-making, then there are a number of issues associated with interpreting and using indicator data that should be explored. Interpretation presents at least four concerns:

- understanding what an indicator value means, what it consists of, and what differences in values it connotes;
- understanding the relationships among variables in an indicator system, including the identification of policy "side-effects";
- attributing causality;
- drawing appropriate policy inferences.

Understanding what an indicator value means

Ensuring that policy-makers and the public understand what an indicator means seems a simpler task than it typically is. Sometimes the difficulty of correcting misunderstandings arises from technical or statistical matters, as when legislators want to require or members of the press want to report that all or most students score at or above the norm on standardized tests. Inaccurate interpretations of indicators are often due to fundamental ignorance of what aggregated data represent, as when scores on college entrance examination tests or on secondary school leaving examinations are compared across jurisdictions or over time, regardless of the fact that the proportions and types of students taking the tests are very different.

Errors such as these can lead to poor policy-making. For example, in the United States, several states and localities have, with no intentional irony, legislated standards including statements like that of New York City's former Comprehensive School

Improvement Plan: "70 per cent of students will score at or above the state reference point (50th percentile)". A number of policies have been enacted on the false belief that norms can be treated as a minimum which all or most students can be expected to meet.

Of course, if one tries hard enough, many things are possible. The furor created by the Cannell Report (1987) – which reported the impossible finding that all American states claimed average student achievement test scores above the national norm – led to concern with the fact that inappropriately used indicators can change behaviour in ways that invalidate the measures themselves (Koretz, 1988). In addition to teaching to the test, these include encouraging poor students to be absent on test days, "helping" students take the test, and outright cheating.

Other measurement issues concern the validity of inferences drawn from indicators. These include concerns of construct validation, data quality, aggregation questions, and a host of other features of data collection and statistical analysis (Guiton and Burstein, 1987; Shavelson *et al.*, 1989). These issues will also have to be faced in other countries that plan to use aggregated test scores as indicators of school quality.

Understanding relationships among indicators

Relationships among variables in a set of indicators affect the inferences drawn from changes in the measures used in developing a construct; so do context changes that influence the meaning of the measures. Guiton and Burstein (1987) offer an example of how reforms aimed at increased course requirements for graduation may actually result in reduced course content or an increase in the drop-out rate. This example demonstrates why indicators are difficult to interpret:

> "In the first instance [dilution of course content], the indicator (number of academic courses) would suggest positive effects due to reform efforts, but achievement scores would not improve (leaving us to question the relationship between coverage and achievement). In the second instance [increases in dropouts], ... improved achievement scores would support the perceived relation between coverage and achievement, but the concomitant reduced student participation would lead to ambiguous interpretations of the indicator (*i.e.* is an increase in the number of courses taken good or bad?)." (Guiton and Burstein, 1987, p. 13)

Attributing causality

Another question that might be asked about this case is whether the increase in average achievement scores actually reflects higher achievement on the part of the individual students remaining in school, or whether it merely reflects a change in the population of students taking the test. By lopping off the lowest scorers, a policy can cause average test scores to increase, even if none of the remaining test-takers improves his score. Often, however, policy-makers will be inclined to infer that if average test scores have increased, the performance of students, indeed the quality of education, has improved. This is an example – one commonly found in the real world of schools and policy – of how inaccurate attributions of causality could lead to an inappropriate policy inference about the success of past efforts and the bases for future courses of action. Indicator systems must be sophisticated enough to anticipate and monitor such policy side-effects.

The resolution of this problem depends in part on the availability of multiple indicators, along with additional studies that investigate the many other variables that may account for changes in either course-taking or achievement. The other key to resolving the problem depends on the training of those who interpret indicator trends for policy-making purposes. Policy-makers must understand not just the meaning of the indicators but also the complex ways in which schools manage the many factors that influence student opportunity and failure. In addition, they must understand how measurements are subject to population shifts and other sources of potential distortion, so that interpretation will lead to policy-making that is at least honest, if not always effective.

Other concerns are also associated with the use of indicators for informing decisions. Even if issues of interpretation are fully resolved, difficulties still remain for determining what policies or practices to change in response to a perceived problem, and how to change the conditions, behaviours, or outcomes that are viewed as problematic. These questions concern the choice of policy targets and strategies.

For example, if indicators suggest that student performance in mathematics is lagging, then policy-makers must consider whether to seek to address the problem by changing policies in any one of many areas: curriculum guidelines or materials, course requirements, test requirements and/or design, teacher education or certification requirements, teacher recruitment policies, school management practices, student grouping or promotion policies, resource allocation policies, or others. Furthermore, decisions must be made about seeking changes at the local school level or another level, depending on views of where either the fault or the more effective lever for change resides.

In addition, there will be decisions about whether to seek change by mandating new policies or practices, by seeking to induce agencies to experiment with new ideas, by creating capacity for change through investment in human resources, by establishing new agencies, or by changing the distribution of decision-making authority (McDonnell and Elmore, 1987).

Drawing appropriate policy inferences

Each of these options suggests a different diagnosis of the source of the problem, and a different theory about what kinds of change might bring about the desired outcomes, while minimising unintended consequences. In order to decide which targets and strategies are worth pursuing, policy-makers must have a great deal of information about how the education system works and how its various components are performing. This suggests, for example, in the case of lagging performance in mathematics, that one would want to examine not only what kind and amount of mathematics is taught to whom, but also who is teaching it, what they know, how well they can meet the intended curriculum goals, and what barriers might prevent them from reaching these goals.

This diagnostic information must then be related to an understanding of the outcomes of prior reform strategies in different circumstances and of the best policy options for the present situation. From a system perspective, one would want to know whether the most productive policy strategy is to mandate a new curriculum, invest in a larger supply of better prepared teachers, change school or student assessment practices, increase the amount of time devoted to mathematics in the schools, equalise resource disparities that produce large differentials in opportunities provided to various groups of children, or some combination of the above.

Fortunately, research studies have offered information pertinent to an understanding of curriculum "inputs" and "processes" that produce educational outcomes. The Second International Mathematics Study, conducted during the early 1980s by the International Association for the Evaluation of Educational Achievement (IEA), for example, examined the type of mathematics taught in the 20 countries involved, in terms of both curricular content and methods. It found few relationships between achievement outcomes and such gross measures of curriculum access as time allocated to mathematics or class size, but considerable links between outcomes and measures of curriculum goals (what students are expected to learn at different grade levels), curriculum intensity (the extent to which certain areas of study are emphasized), curricular organisation (the ways in which topics and skills are structured and sequenced), and curriculum differentiation (the extent to which different expectations are applied to different groups of classes and students) (McKnight *et al.*, 1987).

As one set of indicators, the IEA study provided hypotheses for further exploration. Later, the International Assessment of Educational Progress (IAEP) study confirmed cross-national differences in the extent to which students are exposed to different types of curricular content in science and mathematics (ETS, 1989).

Similarly, analyses of cross-national data show marked differences in course content and student course-taking patterns across countries. For example, in the Second International Mathematics Study, the proportion of 13 year-olds having being taught certain algebra problems ranged from over 90 per cent in France, Hungary, Japan and Thailand to under 40 per cent in New Zealand, Sweden, and Luxembourg, while the proportion of advanced 12th grade students having to master certain areas of calculus ranged from 100 per cent in Japan and Sweden to under 40 per cent in the United States, Thailand and Canadian British Columbia (McKnight *et al.*, 1987, pp. 32-35). Researchers conducting these studies have coined the phrase "opportunity to learn" as a construct for investigating curricular access. Exposure to curriculum content may differ as a consequence of school policies that control course access or teacher decisions that control access to certain knowledge or learning tasks.

Studies have also investigated the ways in which curriculum content is conveyed. For example, the IEA mathematics study obtained detailed information from teachers about their teaching strategies and activities. The results suggest that instruction is dominated by textbooks, with little use of media such as personal computers or calculators and heavy use of "show and tell" approaches for teachers in the United States. The researchers concluded:

> "This use of abstract representations and of strategies geared to rote learning, along with class time devoted to listening to teacher explanations followed by individual seatwork and routine exercises, strongly suggests a view that learning for most students should be passive – teachers transmit knowledge to students who receive it and remember it mostly in the form in which it was transmitted ... In the light of this, it is hardly surprising that the achievement test items on which U.S. students most often showed relatively greater growth were those most suited to performance of rote procedures." (McKnight *et al.*, 1987, p. 81)

In the IAEP study, students were surveyed about their classroom activities, and reported how much time they spent listening to the teacher lecture, doing independent seatwork, reading the text, performing or watching experiments, working in small groups, or working with another classmate (NCES, 1991). These indicators also showed informative differences. In contrast to their peers in other countries, for example, American students reported low levels of involvement in science experimentation and relatively little co-operative group work in mathematics.

To be sure that such observations are defensible, multiple sources of data are needed. Sources of potential bias in the international studies mentioned above, for example, include the sampling design, the extent to which selected schools agree to participate, and differences in the proportions of students in different kinds of educational settings at different ages. Thus, several sources of data are needed to confirm the leads offered by any one study.

With this kind of triangulated information, policy-makers can begin to evaluate the sources of perceived problems. The strategic question (how to make the desired changes) requires still other kinds of knowledge. If curricular changes were sought, what strategies would likely be successful in bringing them about and sustaining them? It has become clear from research studies that, to be successful, policies must take account of and respond to the varying needs and capacities of local agents (Berman and McLaughlin, 1977; Elmore and McLaughlin, 1988). In addition, new policies may find an environment already constrained by prior policies and local conditions that may be hostile to the desired changes (Darling-Hammond, 1990*a*).

A recent investigation of the implementation of a new state mathematics curriculum framework in the United Stated reaffirmed the lessons raised by other change-agent studies (Cohen, 1990). In California, conscientious efforts to introduce a new approach to teaching mathematics (one addressing the kinds of perceived curriculum deficiencies listed above) have not yet succeeded in transforming mathematics teaching in classrooms for two major reasons: first, by and large, teachers have not been prepared in a way that enables them to understand and use the new curricular approach; and second, testing policies undermine rather than reinforce the kind of teaching sought by the new curriculum.

Effective policy strategies will thus need to invest in teacher knowledge as well as in new assessment strategies, if the curriculum goals are to be achieved. This example also illustrates the need for policy implementation studies to supplement an indicator system in order to support interpretation. Their diagnoses may point the way to constructive policy options.

Dangers in over-using indicator data for policy development

The examples above illustrate ways in which a well-designed indicator system, resting on rich sources of data about classrooms, can provide policy-makers with clues as to the relationships between education inputs, processes, and outcomes. As indicator data become more widely available, however, there are at least three potential dangers that should be kept in mind.

First, overzealous efforts to push data beyond their proper limits may yield misleading results. For example, attempts to answer difficult questions about causal relationships, such as programme effects, using statistical methods to partial out the variance components associated with different variables in large, cross-sectional data sets – a request often made by U.S. policy-makers of the National Assessment of Educational Progress (NAEP) data – are inappropriate and likely to produce spurious findings (NCES, 1985).

The attempt to draw inferences where none are truly supportable can have real-world effects beyond the besmirching of scientific canons. Educational programmes, along with hypotheses, may rise or fall in the process.

Second, the mere publication of indicator data can change behaviour in ways that invalidate the indicator. When such data are used in incentive systems, such that decisions are made or rewards or sanctions are distributed based on the data, these effects are exacerbated. Haney and Madaus (1986) have documented the effects on curriculum and teaching of "high stakes" testing programmes – those that make decisions about students or school programmes based on test results – in several countries. The tendency to "teach to the test" under such circumstances can invalidate the inferences made about test score meanings, because the assumption that student performance on the items sampled on the test fairly represents curriculum goals has been violated.

Third, the fact that policy-makers tend to look at indicator data as a basis for making changes in policy occasions other kinds of behavioural responses. Oakes (1986) reports that state-level indicator systems in the United States have produced a variety of pressures to misreport:

> "In some states currently collecting indicator data about school processes, school administrators have been reported to encourage students to 'exaggerate' their responses to particular questions about school experiences. In other areas, informal teacher-networks have spread the word among their colleagues to deflate their salary figures and inflate their teaching-load numbers on state data-collection instruments, reasoning that the resulting data might bring about more favourable policies in both areas. These pressures are likely to be proportionate to what can be lost or gained by indicators." (p. 30)

These examples suggest that some responses may make it difficult for indicator interpreters to know what their statistics represent.

Concerns about comparisons among schools, districts, states, provinces, or nations are born of worries that inaccurate or unfair inferences will be made about the causes or meanings of differences in practices or outcomes, and that these inferences will be used to support decisions that could be harmful to some students or schools. Such concerns prevented the use of NAEP scores to support comparisons among states when that assessment was first created in 1965, but with enthusiasm for indicators currently running high, an experiment with state-level comparisons has recently been launched. While endorsing the comparisons, the Council of Chief State School Officers (CCSSO) has also given voice to its members' misgivings, which echo those aired by educators in several OECD countries:

> "Some are worried that federal, state, and local policy-makers may misuse the data, making inappropriate inferences and drawing unwarranted cause and effect conclusions. Fears are expressed that the test will be very influential, and with that influence, foster a national curriculum. Still others fear that the compromises that might be made on objectives will result in an assessment that measures the least common denominator and discourages badly needed curriculum reform." (CCSSO, 1988, p. 1)

Murnane (1987) speculates that responses to the new state-by-state NAEP comparisons may comprise efforts to exclude low-scoring children from the tested sample, to focus instruction on the skills to be tested, and to teach test-taking skills. He notes that the extent to which "teaching to the test" is problematic depends largely on judgements about the quality of the tests and what they measure.

The burden that policy use places on performance measures

As survey research and test score data are expanded as tools for policy-makers, concerns are being raised about the validity and utility of the measures on which the indicators are based and about the uses to which they are put. These concerns are closely related. It may always be true that in some sense statistics can be made to "lie"; however, this generates little concern, so long as they are not used in ways that affect people's lives. As education indicators are used for a wider variety of purposes – decisions about student placement, promotion, and graduation; arbiters of teacher retention and wages; triggers for public intervention in school affairs – issues of measurement come to the fore. It is this instrumental use of indicators – their direct translation into political and administrative mechanisms of control – which creates a burden that is, arguably, based on an ultimately unsustainable illusion.

One example is the concern that instruction will be driven by limited performance measures. Knowledge about human learning and performance has suggested that many currently used tests (especially those which are norm-referenced and use multiple-choice response formats) fail to adequately measure higher-order cognitive skills and abilities (Resnick, 1987; Sternberg, 1985); and many fear that their influence on instruction will cause these kinds of skills and abilities to be neglected (Haney and Madaus, 1986; National Academy of Education, 1987).

In the United States, low absolute levels in students' higher-order thinking and problem-solving abilities have been attributed by NAEP officials, the National Research Council, and representatives of the National Councils of Teachers of English and Mathematics, among others, to the emphasis on basic skills testing and associated "fill-in-the-blanks teaching" (NAEP, 1979; National Research Council, 1979; Office of Productivity, Technology and Innovation, 1980). This relationship is a function of the instrumental use of test score indicators as the basis for designing and administering managerial sanctions and incentives.

As multiple indicators of the same construct become more widely available, their use can produce pressures to improve the quality of information they provide. In arguing for new indicators of reading education and achievement, Guthrie (1987) aptly described the political rationale for improved measurement:

> "Reading assessments are being deployed by administrators to shape the curriculum ... Tests can be used to set standards and goals, which will influence the content of curricula. Because indicators are not merely passive statistics but are proactive agents of change, they must be selected carefully (p. 1). Even though the search for reading indicators takes place in a policy context which requires measures that are manageable, it is critical to avoid radical over-simplification. The hazard of being simplistic is that Johnny's reading will be misunderstood, teachers will be misled, parents will be outraged, and policy-makers will be frustrated in their attempts to improve reading achievement." (p. 5)

Guthrie argues for measures of reading achievement that "are grounded in research on metacognition and information processing" (p. 21) and that extend beyond current measures of decoding and literal comprehension to include analytic, inferential, and research skills, along with indicators of active reading use. In addition, he argues for indicators of instructional processes that stress not only the quantity of instruction students receive, but also the "quality", *i.e.* "research-based features of exemplary teaching programmes" (p. 18). These include attention to the teaching of cognitive processes, instructional strategies for reflective reading, inferencing, and the use of students' background knowledge.

Researchers in other fields also urge substantial changes in assessment goals and methods. Shavelson *et al.* (1987) argue for tasks that demand more realistic problem-solving skills, using assessment techniques such as clinical interviews (Piaget and Inhelder, 1964) and stimulated recall (Shulman and Elstein, 1975), as well as demonstrations of abilities to construct experiments and find solutions. Similarly, Romberg (1987) argues for new indicators of mathematics achievement that rely on contemporary understandings of cognition and that bring "school mathematics" more in line with the field of mathematics as a discipline.

Clearly, since indicator data are used in policy analysis, great demands are made on construct and content validity as well as on reliability of measurement. These burdens should create pressures for investing in needed improvements and maintaining safeguards in the data collection efforts. The reality, however, is that not all governments are interested in making such investments, and those that are may face both technical and financial obstacles. Therefore, caution in using indicators must continually be urged, and policy-makers must always be presented with many different perspectives on a given issue – including multiple sources of data and multiple interpretations of what these mean.

Finally, serious consideration must be given to how policy-makers may responsibly act on indicator data. Below, two very different scenarios are considered; they embody different perspectives on the role of indicators in policy and, not incidentally, on the meaning of accountability in education.

Using indicators to replace decision-making: example of misuse

In an article on accountability, Linn (1987) poses the following scenarios:

Problem: students are not learning enough. *Solution*: require them to pass a test.

Problem: teacher candidates have poor academic preparation. *Solution*: require them to pass a certification test.

Problem: current teachers are not doing a good job. *Solution*: require them to pass a test for recertification.

Problem: schools are not accountable. *Solution*: require that achievement test results be reported by the individual school (p. 181).

Noting Madaus's (1985) comment that "testing is the darling of policy-makers across the country", Linn writes that the increased emphasis on testing has been accompanied by sanctions associated with test results, with the aim, of course, of improving education:

"No one argues with the merits of this goal, but there is considerable debate about the degree to which the various testing requirements facilitate or impede its realisa-

> tion ... Although it is widely agreed that a test that is valid for one purpose may be quite invalid when used for a different purpose, this distinction is too often ignored." (p. 182)

In his scenarios, Linn cites examples of the use of tests both as policy tools intended to promote change and as indicators of whether change has occurred. This use of indicators as both change agents and measurement tools is often incompatible. Furthermore, the behavioural responses to the indicators are often not the changes policy-makers intended. It frequently happens that policy-makers avoid their obligation to understand why desired performances are not occurring and mandate outcomes without attending to changes in school structures, inputs, or processes. Rather than diagnostic tools, the indicators can become both the content and the measure of the policy.

It also happens that measures initially created as indicators of student performance are being used as the sole arbiters of decisions about students or schools. In these cases, the indicators are not used as clues about performance or even as goals; they have become themselves the decision-makers. This has been the case in some states and local districts in the United States, where policies were enacted requiring that test scores be used as the criterion for decisions about student promotion from one grade to the next. In Georgia, this policy even included kindergarten pupils.

Since such student promotion policies were enacted, a substantial body of research has demonstrated that the effects of automatic decision-making are more negative than positive. When students who were retained in a grade are compared to students of equal achievement levels who were promoted, the students who were retained consistently lag on both achievement and social-emotional measures (Shephard and Smith, 1986; Holmes and Matthews, 1984; Rose *et al.*, 1983). As Shephard and Smith put it:

> "Contrary to popular beliefs, repeating a grade does *not* help students gain ground academically and has a negative impact on social adjustment and self-esteem. Ironically, reviewers have also found that the practice of holding children back does not increase the homogeneity of classrooms." (p. 84)

Furthermore, the practice of retaining students is a major contributor to increased drop-out rates. Research suggests that being retained in a grade increases the likelihood of dropping out by 40 to 50 per cent. A second retention increases the risk by 90 per cent (Mann, 1987; see also Carnegie Council on Adolescent Development, 1989; Wehlage *et al.*, 1990; Massachusetts Advocacy Center, 1988). Thus, the policy of automatically retaining students based on a single indicator – their test score performance – produced lower achievement for these students, lower self-esteem, and higher drop-out rates. To be sure, the same negative outcomes could hold if the students were retained for other reasons; the point is that, once an automatic trigger is embedded in policy, options for other educational decisions are foreclosed.

Fortunately, some policy-makers have begun to take note, and the policies are being reversed in some jurisdictions. A parent revolt in North Carolina caused the repeal of that state's early grades testing requirements in 1988. More recently, the Massachusetts Commissioner of Education urged the state's districts to cease retaining low-achieving students (Rothman, 1990, p. 17). In June 1990, the New York City Schools Chancellor Joseph Fernandez repealed his predecessor's "Promotional Gates" policy, which had mandated the retention of students who scored below test cut-off points at certain grade levels, citing research on the relationship between retention and drop-out rates.

Even as the use of indicator data to make mechanical decisions about student placements or promotions may be waning, policy-makers are considering a similar use of indicator data to make decisions about schools. An analogous policy proposal has been made with respect to schools (and enacted, though not yet implemented, in the state of Kentucky): that all schools which do not meet specified performance standards will automatically suffer sanctions, including actions against staff. Those that meet the standards will be rewarded. In the name of accountability for performance, indicators thus become automatic triggers for specific actions. This means that diagnosis and decision-making can be largely avoided. There is no requirement that anyone carefully examine the reasons for lack of performance or worry about what course of action is appropriate. The indicator system does the entire job.

The Kentucky plan mentioned above would use percentage changes in schools' average achievement test scores and other school-level indicator data (*e.g.* drop-out rates) as the basis for automatic allocations of rewards and punishments. Oblivious to the fact that schools' average scores on any measure are sensitive to changes in the population of students taking the test, and that changes in "school scores" over time are often based on substantially different student populations which may change due to high turnover and to manipulation of enrolments and pupil classifications, these policies will create and sustain a variety of undesirable incentives.

Dysfunctional consequences have already been reported from efforts to use the average test scores of schools for making decisions about rewards and sanctions. These include grouping large numbers of low-scoring students in special education placements, so that their scores do not "count" in school reports; excluding low-scoring students from admission to schools when open enrolments are supposedly available or when transfers are requested; and encouraging such students to drop out when they are old enough. An additional consequence, which can undermine the quality of instruction for all students, occurs when schools opt to "teach to the test" (Darling-Hammond and Wise, 1985; Koretz, 1988; Talmage and Rasher, 1980). These policies will also further exacerbate existing incentives for talented staff to opt for schools where students are easy to teach and stability is high. This will even further compromise the educational chances of disadvantaged students, who are already served by a disproportionate share of inexperienced, unprepared and under-qualified teachers (Darling-Hammond, 1990*b*; Oakes, 1990).

Existing policies do not appear to acknowledge the double jeopardy into which students in low-performing schools are being placed. First, policies that sustain the unequal allocation of resources, including the distribution of high-quality staff and programmes, create low-performing schools which put the students who attend them at risk. Then, policy-makers who choose this course would further penalise these schools and their students by withholding resources or introducing other sanctions. These result both in "blaming the victim" and further deflecting time, energy and attention from solving the basic problems causing poor performance.

These uses of indicators as policy triggers can be seen to undermine rather than enhance genuine accountability, since they shirk responsibility for careful analysis and decision-making in favour of a simplistic, and potentially damaging, "cure". If policy-makers and educators are to be truly accountable for serving students in responsible and responsive ways, then they must use information about educational conditions and progress, along with the best knowledge available about sound educational practice, to evaluate where changes seem warranted and to consider what strategies are likely to be supportive.

Using indicators to inform decision-making: example of wise use

In the final analysis, the measurement questions come down to educational goals and social values. What should we encourage? How can our systems of observation and enumeration help to steer a course towards the chosen ends?

As Murnane (1987) notes, the NAEP achievement trend data raised questions that the study was not designed to answer. Why, for example, did the achievement of black 9 year-old students climb steadily and more quickly than that of white students throughout the 1970s? And why has the performance level of 17 year-olds seemed to decline during the 1980s in many subject areas, especially on tasks requiring critical thinking and problem-solving skills? These trends, first noted in the NAEP, have since been confirmed by analyses of several other data sources (Koretz, 1986), but the range of causes and implications are not yet uncovered.

The availability of data to support these kinds of questions can help focus attention on concerns that reflect social goals and educational values. To begin to evaluate policy alternatives, we also need data that permit analyses of what kinds of students are afforded what types of educational experiences, as well as analyses of whether curricular offerings and teaching strategies vary among schools of different types, across sectors (or nations), and over time.

Oakes (1989) has argued that school context indicators – information about resources, policies, organisational structures and processes – are at least as important as outcome measures for policy-makers seeking to improve education: "Such information is essential if they want monitoring and accountability systems to mirror the condition of education accurately or to be useful for making improvements" (p. 182). It can be added that those who would attempt to use indicators in the quest for accountability and improvement can themselves be held accountable for making sound decisions. If data are not available about what schools are doing, then theories about why performance is sagging or succeeding will be based on each interpreter's ideas about education, rather than on information about what actually is occurring in schools and classrooms.

Accountability cannot be achieved merely by the analysis of performance indicators and their consequences; it also requires attention to monitoring the nature and distribution of schooling opportunities. The task of educators is to discover and adopt practices that serve students well; the task of policy-makers is to create schools and conditions that will promote such practices. Genuine accountability requires that practices and policies continually be evaluated and revised in light of how well they achieve these goals. Indicators that serve the cause of genuine accountability will point the attention of both educators and policy-makers to their primary tasks and will display the extent to which important ingredients of good schooling are available.

In the United States, for example, assessment data from many sources have long shown disparities between non-Asian minority students and white and Asian students on many achievement measures. Alone, however, these data are not helpful for understanding the sources of these disparities. In the absence of other indicators, many would assume that all students had access to equivalent educational opportunities, but that some were failing to take advantage of them, for some unidentified reasons.

However, analyses of participation data from the "High School and Beyond" surveys and several other data sets compiled in the United States reveal dramatic differences among students of various racial/ethnic and social groups in course-taking patterns. These data also show that mathematics course-taking is strongly related to achievement for all groups, so that students with similar course-taking records show virtually no difference by race or ethnicity in achievement test scores. Recent research also demonstrates that the availability of these key courses varies substantially among schools serving different racial and socio-economic student populations (Oakes, 1990).

Furthermore, several recently developed sets of indicators show that both the availability and qualifications of teachers are dramatically different in schools serving different types of students. Students in high-minority schools are much less likely than students in low-minority schools to have access to mathematics and science teachers who are experienced and well-trained in their subject areas (Oakes, 1990; Darling-Hammond, 1990*b*). Thus, indicators which open the "black box" of schooling offer an opportunity for policy-makers to identify and begin to address some of the factors associated with poor performance.

In some indicator systems, school context data are being supplied. Such data help to illuminate the nature and distribution of curricular opportunities available to students, and to suggest how these may be associated with what students learn. The U.S. Council of Chief State School Officers (CCSSO) has begun to collect and disseminate state-by-state indicators of school science and mathematics achievement that include indicators of student course-taking patterns by course type, difficulty level, and student characteristics such as sex and grade level. Information is also supplied about mathematics and science teachers, such as their availability, their employment status and conditions, and their teaching qualifications (CCSSO, 1989).

These data have stimulated some states to evaluate the curricular opportunities made available to their students and to take steps to improve them. Measures include recruiting and training well-prepared mathematics and science teachers, upgrading the quality of curricular materials available to schools, and launching attempts to equalise the distribution of teachers, courses, and other teaching resources across schools.

Various states collect and publish their own indicator data focusing on measures of student participation and school opportunities. New York is among those that publish annual indicator data, including such measures as the proportion of students enrolled in (and successful in) the state's advanced Regents courses and the proportion of teachers in each district who are experienced and certified in their fields. Rhode Island publishes district-level data that describe the proportion of students participating in different types of programmes (vocational, academic, remedial, gifted and talented, etc.), and the allocations of funds both across these programmes and on various school functions (*e.g.* instruction, administration, transportation, and so on).

California provides each school with a detailed report, including participation rates in academic courses, attendance and drop-out rates, indicators of students' social attitudes and perceptions regarding the school and peers, along with mobility data and test score information disaggregated by income group and compared to those of schools with similar student bodies (Archbald and Newmann, 1988). California also encourages school districts to develop and report local indicators and suggests other data (*e.g.* number of writing assignments; amount of homework) that might be reported. High-performing schools are reviewed on a three-year cycle by the state; low-performing schools are provided technical assistance by the state and are reviewed more often (OERI, 1988).

Where these kinds of information are made available, they enter state and local educational decisions in a variety of ways. Their most important contributions to policy-making are, first, that they provide a basis for the diagnosis of perceived problems and, second, that they encourage policy-makers and educators to examine and improve the opportunities made available to students.

Conclusion

The above discussion suggests at least the following five guidelines for the constructive use of indicators in guiding and responding to policy:

Numbers are not enough

Interpretation of indicators by policy-makers, educators and the public must be informed by knowledge about teaching, learning, and the functioning of schools. It is widely known that quantitative data tend, over time, to override consideration of other kinds of information. This is especially problematic, since the meaning and limitations of quantitative indicators and the inferences it is appropriate to draw from them are often poorly understood. As Sternberg (1985) warns: "The appearance of precision is no substitute for the fact of validity." And validity is dependent on the inferences drawn, not just the nature of the measure itself.

Thus, the development of an indicator system should be accompanied by a concerted effort to educate users about the relevance and meaning of particular kinds of information, including the ways in which opportunities and performances are shaped by school policies, resources, practices, and non-school factors. Descriptive qualitative information about school practices should accompany quantitative indicators. Furthermore, a discussion of their meaning, limitations, and interpretations should always accompany the reporting of indicators.

Multiple indicators are needed

Change in a single indicator, such as school attendance, can rarely, if ever, be accurately interpreted without information on related variables such as student population changes or curriculum adjustments. The presentation of single indicators in isolation from a systemic model of educational participation and performance may lead to inaccurate assumptions about what has really changed and inappropriate inferences about policy needs or effects.

Quite often, a focus on isolated indicators leads to a preoccupation with first-order solutions, where policy-makers or educators may try to elicit changes in the variable under scrutiny by treating that issue without examining what other changes would be needed to achieve the underlying goals of interest. Better policy-making is likely to result from a greater array of relevant clues, providing a more comprehensive picture of the situation.

In this sense, it is probably best that those "policy needs" which are keenly felt at any given time should not be the basis for designing an indicator system, because they might limit or skew the availability of data in ways dictated either by a partisan political agenda or the happenstance of currently popular policy theories. Also, the aspects of schooling that are the direct targets of policy are not the only ones that need to be monitored. Informed policies will need to be based on an understanding of how the context of schooling affects policy implementation. Thus, the form of the indicator system ought to be determined by those enduring features of the education system that influence the outcomes of policies, as well as by the features which policies directly seek to affect.

Indicator systems should be both "redundant" and continually revised

One point made in this chapter is that if indicators are used to inform important policy decisions, then individual and organisational behaviours are likely to change to maximise performance on the indicators chosen as proxies for the constructs of interest. As a consequence of this predictable behaviour, one must be prepared for the possibility that the selected indicator will, over time, become a less good measure of the dimension or phenomenon it is expected to represent.

Koretz (1987) shows how different measures of student performance can change over time, depending upon the aspects of learning they measure and how various conditions and incentives influence schools' attention to them. The trade-offs that are inevitably present can be evaluated when multiple measures of a construct are available. The potential for indicator corruption also lessens when multiple measures of a given aspect of the system are available. Interpretation is also improved when indicators are revised to take account of what is learned about the organisational or instructional side-effects of relying on a particular measure.

Policies must be based on a sound understanding of individual and organisational changes in response to performance data

While indicators may provide information about aspects of education systems and student learning, they cannot, by themselves, solve strategic questions of policy-making. Questions concerning the appropriate strategy for addressing root causes of problems or creating new incentives for success require other kinds of information. Knowledge of how the performance of one aspect of the system is dependent on what occurs in other components must be married with an understanding of how organisations change, how they respond to different types of policy instruments under different conditions, and how they respond to performance measures.

The temptation policy-makers and educators often face is to focus narrowly on improving the particular indicator of interest by turning to first-order solutions and mandating a change, rather than seeking to understand how the indicator may reflect influences from other parts of the system. First-order solutions are rarely successful in the long run, and they may create behavioural changes that invalidate the usefulness of the targeted indicator. In particular, strategic thinking must be informed by an understanding of the capacities of different parts of the education system to sustain the desired performance or to produce desired changes. Strategic policy-making should aim to build capacity where needed while providing useful incentives for improvement.

Indicators should be used for further evaluation, not as potentially counterproductive mechanical triggers for action

Indicators should be viewed as an ingredient of reflective policy-making, not as a replacement for thoughtful evaluation and complex decision-making. As Colin Power, at an OECD meeting organised by the INES project has argued, the value of indicators "is that they provide signals of outstanding questions to be resolved or potential problems that can then be subjected to more intensive scrutiny" (OECD, 1987, p. 6). When specific actions automatically result from changes in the ostensible performance of students, schools, or other jurisdictions on particular measures, they can have many dysfunctional consequences. However, responsible use of indicators as a means for asking more informed policy questions can create incentives for both policy-makers and educators to aim towards more responsive schools.

Bibliography

ARCHBALD, D.A. and NEWMAN, F. (1988), *Beyond Standardized Testing: Assessing Authentic Achievement in the Secondary School*, National Association of Secondary School Principals, Reston, VA.

BERMAN, P. and McLAUGHLIN, M.W. (1977), *Federal Programs Supporting Educational Change*, Vol. 7: *Factors Affecting Implementation and Continuation*, The Rand Corporation, Santa Monica, CA.

CALLAHAN, R.E. (1962), *Education and the Cult of Efficiency*, University of Chicago Press, Chicago.

CANNELL, J.J. (1987), *Nationally Normed Achievement Testing in America's Public Schools: How All Fifty States are Above the National Average*, Friends for Education, Daniels, WV.

CARNEGIE COUNCIL ON ADOLESCENT DEVELOPMENT (1989), *Turning Points: Preparing Youth for the 21st Century*, Carnegie Corporation of New York, New York.

COHEN, D. (ed.) (1990), *Educational Evaluation and Policy Analysis*, Vol. 12 (3), pp. 347-353.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS (CCSSO) (1988), *On Reporting Student Achievement at the State Level by the National Assessment of Educational Progress*, Council of Chief State School Officers, Washington, D.C.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS (CCSSO) (1989), *State By State Indicators of Science and Mathematics Education: Preliminary Report*, Council of Chief State School Officers, Washington, D.C.

CREMIN, L. (1961), *The Transformation of the School: Progressivism in American Education, 1876-1957*, Vintage Books, New York.

DARLING-HAMMOND, L. (1990*a*), "Instructional Policy into Practice: The Power of the Bottom Over the Top", *Educational Evaluation and Policy Analysis*, Vol. 12 (3), pp. 233-241.

DARLING-HAMMOND, L. (1990*b*), "Teacher quality and Equality", in Goodlad, J. and Keating, P. (eds.), *Access to Knowledge: An Agenda for Our Nation's Schools*, College Entrance Examination Board, New York.

DARLING-HAMMOND, L. and WISE, A.E. (1985), "Beyond Standardization: State Standards and School Improvement", *The Elementary School Journal*, Vol. 85 (3), pp. 315-336.

DE NEUFVILLE, J.I. (1975), *Social Indicators and Public Policy: Interactive Processes of Design and Application*, Elsevier, New York.

DEWEY, J. (1904), *The Sources of a Science of Education*, University of Chicago Press, Chicago.

EDUCATIONAL TESTING SERVICE (ETS) (1989), *A World of Differences: An International Assessment of Mathematics and Science*, ETS, Princeton, NJ.

EIDE, K. (1987), "The Need for Statistical Indicators in Education", paper presented at the OECD Conference on Educational Indicators, November, Washington, D.C.

EISNER, E.W. and VALLANCE, E. (eds.) (1974), *Conflicting Conceptions of Curriculum*, McCutchan Publishing Corporation, Berkeley, CA.

BEST COPY AVAILABLE

ELMORE, R. and McLAUGHLIN, M.W. (1988), *Steady Work*, The Rand Corporation, Santa Monica, CA.

GUITON, G. and BURSTEIN, L. (1987), *Steps in Construct Validation within a Conceptual Framework for the Educational System*, Center for Research Evaluation, Standards and Student Testing, University of California at Los Angeles, Los Angeles, CA.

GUTHRIE, J. (1987), *Indicators of Reading Achievement*, The Rand Corporation, Santa Monica, CA.

HANEY, W. and MADAUS, G. (1986), "Effects of Standardized Testing and the Future of the National Assessment of Educational Progress", Working paper for the NAEP study group, Center for the Study of Testing, Evaluation and Educational Policy, Chestnut Hill, MA.

HOLMES, C.T. and MATTHEWS, K.M. (1984), "The Effects of Nonpromotion on Elementary and Junior High School Pupils: A Meta-analysis", *Review of Educational Research*, Vol. 54, pp. 225-236.

KORETZ, D. (1986), *Trends in Educational Achievement*, Congressional Budget Office, Washington, D.C.

KORETZ, D. (1987), *Trends in Educational Achievement: Explanations and Implications*, Congressional Budget Office, Washington, D.C.

KORETZ, D. (1988), "Arriving in Lake Wobegon: Are Standardized Tests Exaggerating Achievement and Distorting Instruction?", *American Educator*, Vol. 12 (2), pp. 8-15 and pp. 46-52.

LINN, R.L. (1987), "Accountability: The Comparison of Educational Systems and the Quality of Test Results", *Educational Policy*, Vol. 1 (2), pp. 181-198.

MADAUS, G.F. (1985), "Public Policy and the Testing Profession: You've Never Had it So Good?", *Educational Measurement: Issues and Practice*, Vol. 4 (1), pp. 5-11.

MANN, D. (1987), "Can We Help Dropouts? Thinking about the Undoable", in Natriello, G. (ed.), *School Dropouts: Patterns and Policies*, Teachers College Press, New York.

MASSACHUSETTS ADVOCACY CENTER AND THE CENTER FOR EARLY ADOLESCENCE (1988), *Before It's Too Late: Dropout Prevention in the Middle Grades*, Massachusetts Advocacy Center, Boston.

McDONNELL, L. and ELMORE, R. (1987), *Alternative Policy Instruments*, The Rand Corporation, Santa Monica, CA.

McKNIGHT, C.C., CROSSWHITE, F.J., DOSSEY, J.A., KIFER, E., SWAFFORD, S.O., TRAVERS, K.J. and COONEY, T.J. (1987), *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective*, Stipes Publishing, Champaign, IL.

MURNANE, R. (1987), "Improving Education Indicators and Economic Indicators: The Same Problems?", *Educational Evaluation and Policy Analysis*, Vol. 9 (2), pp. 101-116.

NATIONAL ACADEMY OF EDUCATION (1987), *The Nation's Report Card*, National Academy of Education, Cambridge, MA.

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP) (1979), *Changes in Mathematical Achievement, 1973-78*, NAEP, Denver, CO.

NATIONAL CENTER FOR EDUCATION STATISTICS (NCES) (1985), *Synthesis of Invited Papers: Elementary/Secondary Education Data Redesign Project*, National Center for Education Statistics, U.S. Department of Education, Washington, D.C.

NATIONAL CENTER FOR EDUCATION STATISTICS (NCES) (1991), *The State of Mathematics Achievement: NAEP's 1990 Assessment of the National and the Trial Assessment of the States*, National Center for Education Statistics, U.S. Department of Education, Washington, D.C.

NATIONAL RESEARCH COUNCIL (1979), *The State of School Science*, National Research Council, Commission on Human Resources, Washington, D.C.

BEST COPY AVAILABLE

OAKES, J. (1985), *Keeping Track: How Schools Structure Inequality*, Yale University Press, New Haven, CT.

OAKES, J. (1986), *Educational Indicators: A Guide for Policymakers*, The Rand Corporation, Santa Monica, CA.

OAKES, J. (1989), "What Educational Indicators? The Case for Assessing the School Context", *Educational Evaluation and Policy Analysis*, Vol. 11 (2), pp. 181-199.

OAKES, J. (1990), *Multiplying Inequalities*, The Rand Corporation, Santa Monica, CA.

OECD (1987), "International Conference on Educational Indicators: Conclusions and Follow-up", November, OECD, Paris (mimeo).

OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT (OERI) (1988), *Creating Responsible and Responsive Accountability Systems*, U.S. Department of Education, Washington, D.C.

OFFICE OF PRODUCTIVITY, TECHNOLOGY, AND INNOVATION (1980), *Learning Environments for Innovation*, Department of Commerce, Washington, D.C.

PIAGET, J. and INHELDER, B. (1964), *The Early Growth of Logic in the Child*, Routledge and Kegan Paul, London.

RESNICK, L.B. (1987), *Education and Learning to Think*, National Academy Press, Washington, D.C.

ROMBERG, T.A. (1987), "School Mathematics Indicators", paper prepared for the AERA Annual Meeting, Washington, D.C.

ROSE, J.S., MEDWAY, F.J., CANTRELL, V.L. and MARUS, S.H. (1983), "A Fresh Look at the Retention-Promotion Controversy", *Journal of School Psychology*, Vol. 21, pp. 201-211.

ROTHMAN, R. (1990), "Abolish Grade Retention, Massachusetts Chief Says", *Education Week*, 2 May, p. 17.

RUBY, A. (1989), *Indicators Resource Folder*, Department of Education, New South Wales.

SHAVELSON, R.J., McDONNELL, L.M. and OAKES, J. (eds.) (1989), *Indicators for Monitoring Mathematics and Science Education: A Sourcebook*, The Rand Corporation, Santa Monica, CA.

SHAVELSON, R.J., McDONNELL, L.M., OAKES, J. and CAREY, J. (1987), *Indicator Systems for Monitoring Mathematics and Science Education*, The Rand Corporation, Santa Monica, CA.

SHELDON, E.B. and PARKE, R. (1975), "Social Indicators" *Science*, Vol. 188, pp. 693-699.

SHEPHARD, L.A. and SMITH, M.L. (1986), "Synthesis of Research on School Readiness and Kindergarten Retention", *Educational Leadership*, November, pp. 78-86.

SHULMAN, L. and ELSTEIN, A. (1975), "Studies of Problem Solving, Judgement, and Decision Making: Implications for Educational Research", in Kerlinger, F. (ed.), *Review of Research in Education*, Peacock, Itasca, IL.

STERNBERG, R.J. (1985), *Beyond IQ*, Cambridge University Press, New York.

TALMAGE, H. and RASHER, S.P. (1980), "Unanticipated Outcomes: The Perils to Curriculum Goals", *Phi Delta Kappan*, November, pp. 194-198.

TYACK, D. (1974), *The One Best System*, Harvard University Press, Cambridge, MA.

WEHLAGE, G.G., ROBERT, R.A., RUTTER, A., SMITH, G.A., LESKO, N. and FERNANDEZ, R.R. (1990), *Reducing the Risk: Schools as Communities of Support*, Falmer Press, New York.

WISE, A.E. and DARLING-HAMMOND, L. (1984), "Education by Voucher: Private Choice and the Public Interest", *Educational Theory*, Vol. 34 (1), pp. 29-47.

BEST COPY AVAILABLE

MAIN SALES OUTLETS OF OECD PUBLICATIONS
PRINCIPAUX POINTS DE VENTE DES PUBLICATIONS DE L'OCDE

ARGENTINA – ARGENTINE
Carlos Hirsch S.R.L.
Galería Güemes, Florida 165, 4° Piso
1333 Buenos Aires Tel. (1) 331.1787 y 331.2391
Telefax: (1) 331.1787

AUSTRALIA – AUSTRALIE
D.A. Information Services
648 Whitehorse Road, P.O.B 163
Mitcham, Victoria 3132 Tel. (03) 873.4411
Telefax: (03) 873.5679

AUSTRIA – AUTRICHE
Gerold & Co.
Graben 31
Wien I Tel. (0222) 533.50.14

BELGIUM – BELGIQUE
Jean De Lannoy
Avenue du Roi 202
B-1060 Bruxelles Tel. (02) 538.51.69/538.08.41
Telefax: (02) 538.08.41

CANADA
Renouf Publishing Company Ltd.
1294 Algoma Road
Ottawa, ON K1B 3W8 Tel. (613) 741.4333
Telefax: (613) 741.5439

Stores:
61 Sparks Street
Ottawa, ON K1P 5R1 Tel. (613) 238.8985
211 Yonge Street
Toronto, ON M5B 1M4 Tel. (416) 363.3171
Telefax: (416)363.59.63

Les Éditions La Liberté Inc.
3020 Chemin Sainte-Foy
Sainte-Foy, PQ G1X 3V6 Tel. (418) 658.3763
Telefax: (418) 658.3763

Federal Publications Inc.
165 University Avenue, Suite 701
Toronto, ON M5H 3B8 Tel. (416) 860.1611
Telefax: (416) 860.1608

Les Publications Fédérales
1185 Université
Montréal, QC H3B 3A7 Tel. (514) 954.1633
Telefax : (514) 954.1635

CHINA – CHINE
China National Publications Import
Export Corporation (CNPIEC)
16 Gongti E. Road, Chaoyang District
P.O. Box 88 or 50
Beijing 100704 PR Tel. (01) 506.6688
Telefax: (01) 506.3101

DENMARK – DANEMARK
Munksgaard Book and Subscription Service
35, Nørre Søgade, P.O. Box 2148
DK-1016 København K Tel. (33) 12.85.70
Telefax: (33) 12.93.87

FINLAND – FINLANDE
Akateeminen Kirjakauppa
Keskuskatu 1, P.O. Box 128
00100 Helsinki

Subscription Services/Agence d'abonnements :
P.O. Box 23
00371 Helsinki Tel. (358 0) 12141
Telefax: (358 0) 121.4450

FRANCE
OECD/OCDE
Mail Orders/Commandes par correspondance:
2, rue André-Pascal
75775 Paris Cedex 16 Tel. (33-1) 45.24.82.00
Telefax: (33-1) 49.10.42.76
Telex: 640048 OCDE

Orders via Minitel, France only/
Commandes par Minitel, France exclusivement :
36 15 OCDE

OECD Bookshop/Librairie de l'OCDE :
33, rue Octave-Feuillet
75016 Paris Tel. (33-1) 45.24.81.67
(33-1) 45.24.81.81

Documentation Française
29, quai Voltaire
75007 Paris Tel. 40.15.70.00

Gibert Jeune (Droit-Économie)
6, place Saint-Michel
75006 Paris Tel. 43.25.91.19

Librairie du Commerce International
10, avenue d'Iéna
75016 Paris Tel. 40.73.34.60

Librairie Dunod
Université Paris-Dauphine
Place du Maréchal de Lattre de Tassigny
75016 Paris Tel. (1) 44.05.40.13

Librairie Lavoisier
11, rue Lavoisier
75008 Paris Tel. 42.65.39.95

Librairie L.G.D.J. - Montchrestien
20, rue Soufflot
75005 Paris Tel. 46.33.89.85

Librairie des Sciences Politiques
30, rue Saint-Guillaume
75007 Paris Tel. 45.48.36.02

P.U.F.
49, boulevard Saint-Michel
75005 Paris Tel. 43.25.83.40

Librairie de l'Université
12a, rue Nazareth
13100 Aix-en-Provence Tel. (16) 42.26.18.08

Documentation Française
165, rue Garibaldi
69003 Lyon Tel. (16) 78.63.32.23

Librairie Decitre
29, place Bellecour
69002 Lyon Tel. (16) 72.40.54.54

GERMANY – ALLEMAGNE
OECD Publications and Information Centre
August-Bebel-Allee 6
D-53175 Bonn Tel. (0228) 959.120
Telefax: (0228) 959.12.17

GREECE – GRÈCE
Librairie Kauffmann
Mavrokordatou 9
106 78 Athens Tel. (01) 32.55.321
Telefax: (01) 36.33.967

HONG-KONG
Swindon Book Co. Ltd.
13–15 Lock Road
Kowloon, Hong Kong Tel. 366.80.31
Telefax: 739.49.75

HUNGARY – HONGRIE
Euro Info Service
Margitsziget, Európa Ház
1138 Budapest Tel. (1) 111.62.16
Telefax : (1) 111.60.61

ICELAND – ISLANDE
Mál Mog Menning
Laugavegi 18, Pósthólf 392
121 Reykjavik Tel. 162.35.23

INDIA – INDE
Oxford Book and Stationery Co.
Scindia House
New Delhi 110001 Tel.(11) 331.5896/5308
Telefax: (11) 332.5993

17 Park Street
Calcutta 700016 Tel. 240832

INDONESIA – INDONÉSIE
Pdii-Lipi
P.O. Box 269/JKSMG/88
Jakarta 12790 Tel. 583467
Telex: 62 875

ISRAEL
Praedicta
5 Shatna Street
P.O. Box 34030
Jerusalem 91430 Tel. (2) 52.84.90/1/2
Telefax: (2) 52.84.93

R.O.Y.
P.O. Box 13056
Tel Aviv 61130 Tél. (3) 49.61.08
Telefax (3) 544.60.39

ITALY – ITALIE
Libreria Commissionaria Sansoni
Via Duca di Calabria 1/1
50125 Firenze Tel. (055) 64.54.15
Telefax: (055) 64.12.57

Via Bartolini 29
20155 Milano Tel. (02) 36.50.83

Editrice e Libreria Herder
Piazza Montecitorio 120
00186 Roma Tel. 679.46.28
Telefax: 678.47.51

Libreria Hoepli
Via Hoepli 5
20121 Milano Tel. (02) 86.54.46
Telefax: (02) 805.28.86

Libreria Scientifica
Dott. Lucio de Biasio 'Aeiou'
Via Coronelli, 6
20146 Milano Tel. (02) 48.95.45.52
Telefax: (02) 48.95.45.48

JAPAN – JAPON
OECD Publications and Information Centre
Landic Akasaka Building
2-3-4 Akasaka, Minato-ku
Tokyo 107 Tel. (81.3) 3586.2016
Telefax: (81.3) 3584.7929

KOREA – CORÉE
Kyobo Book Centre Co. Ltd.
P.O. Box 1658, Kwang Hwa Moon
Seoul Tel. 730.78.91
Telefax: 735.00.30

MALAYSIA – MALAISIE
Co-operative Bookshop Ltd.
University of Malaya
P.O. Box 1127, Jalan Pantai Baru
59700 Kuala Lumpur
Malaysia Tel. 756.5000/756.5425
Telefax: 757.3661

MEXICO – MEXIQUE
Revistas y Periodicos Internacionales S.A. de C.V.
Florencia 57 - 1004
Mexico, D.F. 06600 Tel. 207.81.00
Telefax : 208.39.79

NETHERLANDS – PAYS-BAS
SDU Uitgeverij Plantijnstraat
Externe Fondsen
Postbus 20014
2500 EA's-Gravenhage Tel. (070) 37.89.880
Voor bestellingen: Telefax: (070) 34.75.778

BEST COPY AVAILABLE

NEW ZEALAND
NOUVELLE-ZÉLANDE
Legislation Services
P.O. Box 12418
Thorndon, Wellington Tel. (04) 496.5652
Telefax: (04) 496.5698

NORWAY – NORVÈGE
Narvesen Info Center – NIC
Bertrand Narvesens vei 2
P.O. Box 6125 Etterstad
0602 Oslo 6 Tel. (022) 57.33.00
Telefax: (022) 68.19.01

PAKISTAN
Mirza Book Agency
65 Shahrah Quaid-E-Azam
Lahore 54000 Tel. (42) 353.601
Telefax: (42) 231.730

PHILIPPINE – PHILIPPINES
International Book Center
5th Floor, Filipinas Life Bldg.
Ayala Avenue
Metro Manila Tel. 81.96.76
Telex 23312 RHP PH

PORTUGAL
Livraria Portugal
Rua do Carmo 70-74
Apart. 2681
1200 Lisboa Tel.: (01) 347.49.82/5
Telefax: (01) 347.02.64

SINGAPORE – SINGAPOUR
Gower Asia Pacific Pte Ltd.
Golden Wheel Building
41, Kallang Pudding Road, No. 04-03
Singapore 1334 Tel. 741.5166
Telefax: 742.9356

SPAIN – ESPAGNE
Mundi-Prensa Libros S.A.
Castelló 37, Apartado 1223
Madrid 28001 Tel. (91) 431.33.99
Telefax: (91) 575.39.98

Libreria Internacional AEDOS
Consejo de Ciento 391
08009 – Barcelona Tel. (93) 488.30.09
Telefax: (93) 487.76.59

Llibreria de la Generalitat
Palau Moja
Rambla dels Estudis, 118
08002 – Barcelona
(Subscripcions) Tel. (93) 318.80.12
(Publicacions) Tel. (93) 302.67.23
Telefax: (93) 412.18.54

SRI LANKA
Centre for Policy Research
c/o Colombo Agencies Ltd.
No. 300-304, Galle Road
Colombo 3 Tel. (1) 574240, 573551-2
Telefax: (1) 575394, 510711

SWEDEN – SUÈDE
Fritzes Information Center
Box 16356
Regeringsgatan 12
106 47 Stockholm Tel. (08) 690.90.90
Telefax: (08) 20.50.21

Subscription Agency/Agence d'abonnements :
Wennergren-Williams Info AB
P.O. Box 1305
171 25 Solna Tel. (08) 705.97.50
Téléfax : (08) 27.00.71

SWITZERLAND – SUISSE
Maditec S.A. (Books and Periodicals - Livres et périodiques)
Chemin des Palettes 4
Case postale 266
1020 Renens Tel. (021) 635.08.65
Telefax: (021) 635.07.80

Librairie Payot S.A.
4, place Pépinet
CP 3212
1002 Lausanne Tel. (021) 341.33.48
Telefax: (021) 341.33.45

Librairie Unilivres
6, rue de Candolle
1205 Genève Tel. (022) 320.26.23
Telefax: (022) 329.73.18

Subscription Agency/Agence d'abonnements :
Dynapresse Marketing S.A.
38 avenue Vibert
1227 Carouge Tel.: (022) 308.07.89
Telefax : (022) 308.07.99

See also – Voir aussi :
OECD Publications and Information Centre
August-Bebel-Allee 6
D-53175 Bonn (Germany) Tel. (0228) 959.120
Telefax: (0228) 959.12.17

TAIWAN – FORMOSE
Good Faith Worldwide Int'l. Co. Ltd.
9th Floor, No. 118, Sec. 2
Chung Hsiao E. Road
Taipei Tel. (02) 391.7396/391.7397
Telefax: (02) 394.9176

THAILAND – THAÏLANDE
Suksit Siam Co. Ltd.
113, 115 Fuang Nakhon Rd.
Opp. Wat Rajbopith
Bangkok 10200 Tel. (662) 225.9531/2
Telefax: (662) 222.5188

TURKEY – TURQUIE
Kültür Yayinlari Is-Türk Ltd. Sti.
Atatürk Bulvari No. 191/Kat 13
Kavaklidere/Ankara Tel. 428.11.40 Ext. 2458
Dolmabahce Cad. No. 29
Besiktas/Istanbul Tel. 260.71.88
Telex: 43482B

UNITED KINGDOM – ROYAUME-UNI
HMSO
Gen. enquiries Tel. (071) 873 0011
Postal orders only:
P.O. Box 276, London SW8 5DT
Personal Callers HMSO Bookshop
49 High Holborn, London WC1V 6HB
Telefax: (071) 873 8200
Branches at: Belfast, Birmingham, Bristol, Edinburgh, Manchester

UNITED STATES – ÉTATS-UNIS
OECD Publications and Information Centre
2001 L Street N.W., Suite 700
Washington, D.C. 20036-4910 Tel. (202) 785.6323
Telefax: (202) 785.0350

VENEZUELA
Libreria del Este
Avda F. Miranda 52, Aptdo. 60337
Edificio Galipán
Caracas 106 Tel. 951.1705/951.2307/951.1297
Telegram: Libreste Caracas

Subscription to OECD periodicals may also be placed through main subscription agencies.

Les abonnements aux publications périodiques de l'OCDE peuvent être souscrits auprès des principales agences d'abonnement.

Orders and inquiries from countries where Distributors have not yet been appointed should be sent to: OECD Publications Service, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

Les commandes provenant de pays où l'OCDE n'a pas encore désigné de distributeur peuvent être adressées à : OCDE, Service des Publications, 2, rue André-Pascal, 75775 Paris Cedex 16, France.

11-1994

OECD PUBLICATIONS, 2 rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(96 93 02 1) ISBN 92-64-14050-6 - No. 46785 1994

BEST COPY AVAILABLE

Making Education Count

Developing and Using International Indicators

At a time of general and growing concern about the substance, effectiveness, and cost of education, guidelines for its evaluation are essential for understanding and policy. This volume presents what is currently known about the organisation, development, measurement and uses of international education indicators. The political contexts within which these indicators are used for informing policy-makers receive particular attention. The chapters are thematically grouped to address conceptual and analytical issues in the development and implementation of different types of indicators, including indicators of learning, student achievement and particular educational outcomes such as labour market destinations, as well as the uses and abuses of reporting and interpreting international education indicators.

3 347370 467857

(96 93 02 1) DCX
ISBN 92-64-14050-6

BEST COPY AVAILABLE